

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:21 ; Search time 31.61 Seconds  
(without alignments)  
1633.955 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQVFALCAALLAG.....FRAGLRHGDEILAVRASPRQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802:\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2353	100.0	465	21	AA83150
2	2317	98.5	499	21	AA83150
3	2254	95.8	499	21	AA83150
4	2248	95.5	499	21	AA83150
5	2244	95.4	499	21	AA83150
6	2243.5	95.0	498	21	AA83150
7	2202	93.6	499	21	AA83150
8	2197	93.4	499	21	AA83150
9	2130.5	90.5	475	21	AA83150
10	2092	88.9	414	21	AA83150
11	2002	85.1	448	21	AA83150

12	1832	77.9	370	21	AA83150
13	962.5	40.9	460	20	AA83150
14	731.5	31.1	503	17	AA83150
15	705	30.0	474	22	AA83150
16	685	29.1	491	20	AA83150
17	633.5	26.9	475	12	AA83150
18	632	26.9	463	17	AA83150
19	629	26.7	437	21	AA83150
20	627.5	26.7	455	22	AA83150
21	624	26.5	437	21	AA83150
22	624	26.5	437	21	AA83150
23	623	26.5	437	21	AA83150
24	622	26.4	437	21	AA83150
25	603.5	25.6	491	19	AA83150
26	573.5	24.4	443	20	AA83150
27	565.5	24.0	451	20	AA83150
28	559.5	23.8	387	19	AA83150
29	542.5	23.1	433	21	AA83150
30	542.5	23.1	433	21	AA83150
31	533	22.7	433	21	AA83150
32	449.5	19.3	549	22	AA83150
33	449.5	19.1	397	21	AA83150
34	448	19.0	320	21	AA83150
35	448	19.0	362	21	AA83150
36	448	19.0	394	21	AA83150
37	448	19.0	451	21	AA83150
38	448	19.0	480	18	AA83150
39	448	19.0	480	21	AA83150
40	448	19.0	480	21	AA83150
41	447.5	19.0	355	22	AA83150
42	445	18.9	480	21	AA83150
43	443.5	18.8	405	21	AA83150
44	442.5	18.8	413	22	AA83150
45	442.5	18.8	441	22	AA83150

ALIGNMENTS

RESULT 1  
AA83150  
ID AA83150 standard; Protein; 465 AA.  
XX  
AC AA83150;  
XX  
DT 24-JUL-2000 (first entry)  
XX  
DE NGSP polypeptide of Neisseria gonorrhoeae.  
XX  
KW NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
KW cellular matrix; adherence domain; ligand; detection; diagnosis;  
KW screening; probe; primer; prophylaxis; therapy.  
XX  
OS Neisseria gonorrhoeae.  
XX  
PN WO200012133-A1.  
XX  
PD 09-MAR-2000.  
XX  
PF 01-SEP-1999; 99WO-US20070.  
XX  
PR 01-SEP-1998; 98US-0098685.  
XX  
PA (ANTE-) ANTEX BIOLOGICS INC.  
XX  
PI Jackson WJ, Harris AM;  
XX  
DR WPI: 2000-237782/20.  
XX  
N-PSDB; AA93414.  
PT Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
PT Neisseria useful for diagnosis, prevention or treatment of Neisseria  
PT infections

Neisseria meningit  
Protein encoded by  
Heat shock protein  
Lawsonia intracell  
Protein involved i  
S. typhimurium Htr  
H. influenzae SB33  
Haemophilus influe  
E. coli growth and  
Haemophilus influe  
Haemophilus influe  
Haemophilus influe  
Escherichia coli s  
Antigen 1 from clu  
Protein encoded by  
H. pylori GHPO 536  
BAS011 protein se  
BAS011 protein se  
Mycobacterium tube  
Streptococcus pneu  
Human HtrA (high t  
Human HtrA (high t  
C-terminal domain  
Human HtrA (high t  
Osteoblast like ce  
Human HtrA (high t  
Human serine prote  
E. coli growth and  
Human HtrA (high t  
Breast and ovarian  
C glutamicum prote  
Corynebacterium g1

XX

Claim 5; Page 61-62; 68pp; English.

The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and Arg-Gly-Asn motifs near the C-terminus which function as adherence domains for extracellular matrix proteins. Using the NGSP polypeptide as a vaccine produces antibodies which inhibit binding of N. gonorrhoeae to the host's cellular matrix reducing attachment and/or subsequent invasion. The NGSP polypeptide and its peptide fragments can be used to immunise an animal and produce an immune response. They can also be used as ligands to detect antibodies elicited in response to Neisseria infections and also as antigens or immunogens for inducing Neisseria-specific antibodies which are useful in immunoassays to detect Neisseria in biological specimens. Nucleotides encoding NGSP or its fragments can be used as probes to identify Neisseria in biological specimens by hybridization or polymerase chain reaction amplification. The NGSP polypeptide can also be used in screening assays to identify agents and compounds which useful as diagnostic, prophylactic or therapeutic agents against Neisseria infection.

Sequence 465 AA;

Query Match 100.0%; Score 2353; DB 21; Length 465;  
Best Local Similarity 100.0%; Pred. No. 1.4e-190;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFKKYQYFALAALCAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60  
DB 1 VFKKYQYFALAALCAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60  
QY 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMNPEIQEEADD 120  
DB 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMNPEIQEEADD 120  
QY 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYTKAKLGSDVQSDVALLKIDA 180  
DB 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYTKAKLGSDVQSDVALLKIDA 180  
QY 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240  
DB 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240  
QY 241 INPGNSGGFLPNLKGQVINGSOIYRSRGGFMGISAIPDIDVAMNVAEQLKNTGKVORGO 300  
DB 241 INPGNSGGFLPNLKGQVINGSOIYRSRGGFMGISAIPDIDVAMNVAEQLKNTGKVORGO 300  
QY 301 LGVITQEVSYGLAQSGFLDKAGLAKILPGSPAERAGLQAGDIVLSLDGGETRSSGDL 360  
DB 301 LGVITQEVSYGLAQSGFLDKAGLAKILPGSPAERAGLQAGDIVLSLDGGETRSSGDL 360  
QY 361 PYWGAITPGKEVSLGVWRKGEEITIKALGNAAHTGASSKTDAPYTBQSGTFSVES 420  
DB 361 PYWGAITPGKEVSLGVWRKGEEITIKALGNAAHTGASSKTDAPYTBQSGTFSVES 420  
QY 421 AGITLQTHDSSGKHLVVVRVSDAAERAGLRHGDDEILAVRASPRQ 465  
DB 421 AGITLQTHDSSGKHLVVVRVSDAAERAGLRHGDDEILAVRASPRQ 465

RESULT 2

ID AAY75748 standard; Protein; 499 AA.  
XX AAY75748;  
XX AAY75748;  
DT 21-MAR-2000 (first entry)  
XX Neisseria gonorrhoeae ORF 986 protein sequence SEQ ID NO:2968.  
DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

KW

antibacterial; gene therapy.  
XX Neisseria gonorrhoeae.

OS WO957280-A2.

PN 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

PD 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR ) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizza M, Rappuoli R, Rattl G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

DR N-PSDB; AAZ54510.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

XX Claim 2; Page 1390; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent

CC PCR primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of

CC the invention can be used as vaccines, as diagnostic reagents, and as

CC immunogenic compositions. The polypeptides can be used in the

CC manufacture of medicaments for treating or preventing infection due to

CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the

CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves

CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.

XX Sequence 499 AA;

SQ

Query Match 98.5%; Score 2317; DB 21; Length 499;  
Best Local Similarity 99.1%; Pred. No. 1.7e-187;  
Matches 459; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VFKKYQYFALAALCAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60  
DB 1 VFKKYQYFALAALCAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60  
QY 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMNPEIQEEADD 120  
DB 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMNPEIQEEADD 120  
QY 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYTKAKLGSDVQSDVALLKIDA 180  
DB 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYTKAKLGSDVQSDVALLKIDA 180  
QY 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240  
DB 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240  
QY 241 INPGNSGGFLPNLKGQVINGSOIYRSRGGFMGISAIPDIDVAMNVAEQLKNTGKVORGO 300  
DB 241 INPGNSGGFLPNLKGQVINGSOIYRSRGGFMGISAIPDIDVAMNVAEQLKNTGKVORGO 300

Db 241 inpgnsgplfnlkqgvvinsqlysrsgfmgisfaipidvamnvaeklntgkvqrqg 300  
QY 301 LGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDITVLISLDGGEIRSSGDL 360  
Db 301 lgvitqevsyglaqsfldkasgaliaakiilpgspaeraglaqgdvdivldggeirssgdl 360  
QY 361 PMVVGAITPGKEVSLGVRKGEETIRAKLGNAAEHTGASSKTDDEAPYTEQSGTFSVES 420  
Db 361 pvmvgaitpgkevsylgvrkgeetiraklgnaaehtgassktddeapYTEQSGTFSVES 420  
QY 421 AGITLQHTDSSGKHLVVVRVSDAAERAGLRHGDEILAVRASP 463  
Db 421 agitlqhtdssgkhlvvvrsvdaaeraglrhgdeilavrgvqp 463

## RESULT 3

AAV52995  
ID AAV52995 standard; Protein; 499 AA.

XX AC AAV52995;

XX 21-FEB-2000 (first entry)

XX Neisseria meningitidis strain H44/76 BASB013 protein sequence.

XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial.

XX Neisseria meningitidis.

XX W0955872-A1.

XX 04-NOV-1999.

XX 20-APR-1999; 99WO-EP02765.

XX 23-APR-1998; 98GB-0008734.

XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.

PI Ruelle J;

XX WPI; 2000-052809/04.

DR N-PSDB; AAZ33307.

XX Novel polynucleotides and polypeptides from Neisseria meningitis used  
PT to prepare vaccines against bacterial infections

XX Claim 3; Page 79-81; 94pp; English.

XX The present sequence represents a BASB013 polypeptide isolated from  
XX Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be  
XX employed as research reagents and material for the discovery of  
XX treatments and diagnostics for diseases, particularly human diseases.  
XX They can be used for diagnosis of disease, staging of disease, or  
XX determining response of an infectious organism to drugs. The  
XX polynucleotides may be used as a source for hybridisation probes, and  
XX for screening of genetic mutations, serotype, organism or strain  
XX identification, identification of mutation in BASB013 sequences, and as  
XX components of arrays which are useful for diagnostic and prognostic  
XX purposes. The polypeptides can be used to produce antibodies. The  
XX polypeptides can also be used in vaccine formulations, and to identify  
XX agonists and antagonists. The polypeptides, antibodies, agonists and  
XX antagonists (which are bacteriostatic) are used for the treatment and  
XX prevention of diseases such as upper respiratory tract infection,  
XX invasive bacterial diseases such as bacteraemia and meningitis, and for  
XX the development and screening of antibacterial drugs. They are also used  
XX in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
XX on in-dwelling devices, or to extracellular proteins on wounds, and to  
XX thus prevent tissue damage and/or block the normal progression of  
XX pathogenesis in infections initiated other than by the implantation of  
XX in-dwelling devices or by other surgical techniques.

XX Sequence 499 AA;  
SQ  
Query Match 95.8%; Score 2254; DB 21; Length 499;  
Best Local Similarity 96.1%; Pred. No. 3.6e-182;  
Matches 445; Conservative 6; Mismatches 12; Indels 0; Gaps 0;  
QY 1 VFKKYQYFALAALACALLAGCEKAGSFFGADKKEASFVERIEHTKDGVSMLLPDFAQL 60  
Db 1 mfkkyqylalalcaaslagcdkagsffgadkkeasfveriehtkddgsvsmllpdfaql 60  
QY 61 VOSEGPVVVNIQAAPARTONGSNAETSDPLADSDPVEFFKRLVPMNPEIPQEEADD 120  
Db 61 vqsegpavvniqaapartqngsnaendsdpladndpriefkrlvpmnpeipqeeadd 120  
QY 121 GGLNFGSGPIISKNGYILNTHVVGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180  
Db 121 gglnfgsgfllskdgyilnthvvtgmsikvllndkreytakligsdvqsdvallkida 180  
QY 181 TEELPVVKIGNPNKLPGEWVAAGAPFGFDNSVTAGIVSAKGRSLPNESYTPPIQTDA 240  
Db 181 teelpvkvignpnkdlkpgewvaagapfgfdnsvtagivsakgrslpnesytpfiqtdda 240  
QY 241 INPGNSGGPLFNKLGQVGVINSQIYSRSGFMGIFSFAIPIDVAMNVAEQLKNTGKVORGO 300  
Db 241 inpgnsgplfnlkqgvvinsqlysrsgfmgisfaipidvamnvaeklntgkvqrqg 300  
QY 301 LGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDITVLISLDGGEIRSSGDL 360  
Db 301 lgvitqevsyglaqsfldkasgaliaakiilpgspaeraglaqgdvdivldggeirssgdl 360  
QY 361 PMVVGAITPGKEVSLGVRKGEETIRAKLGNAAEHTGASSKTDDEAPYTEQSGTFSVES 420  
Db 361 pvmvgaitpgkevsylgvrkgeetiraklgnaaehtgassktddeapYTEQSGTFSVES 420  
QY 421 AGITLQHTDSSGKHLVVVRVSDAAERAGLRHGDEILAVRASP 463  
Db 421 agitlqhtdssgkhlvvvrsvdaaeraglrhgdeilavrgvqp 463

## RESULT 4

AAV5749

ID AAV5749 standard; Protein; 499 AA.

XX AC AAV5749;

XX 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2970.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy.

XX Neisseria meningitidis.

XX W09957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099082.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR ) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.  
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX WPI: 2000-062150/05.  
 DR N-PSDB; AA254511.  
 XX Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics -  
 PT Claim 2; Page 1390; 1453pp; English.  
 PS  
 XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 XX Sequence 499 AA;  
 SQ  
 \* Query Match 95.5%; Score 2248; DB 21; Length 499;  
 Best Local Similarity 96.1%; Pred. No. 1.2e-181;  
 Matches 445; Conservative 5; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 VFKKYQYFALAAALCAGCERAGSFFGADKKREASVERIEHTKDDGSVSMILPDPFAQL 60  
 Db 1 vfkkgyalaaalcaaslagckagsgffvadkkaesveriehkddgsvsmilpdtfaql 60  
 QY 61 VQSEGPVAVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMPEIQEEADD 120  
 Db 61 vqsegpavvniqaaprtqngsgnaetdsdpladndpfyefkrlvpmpeiqeeadd 120  
 QY 121 GGLNFGSGFIISKNGYILTNTHVVMGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180  
 Db 121 gglnfgsgfiiskngyiltnthvvtgmgsikvllndkreytakligsdvqsdvallkida 180  
 QY 181 TEELPVVKIGNPKNLGPEWVAIGAPFGPDNSVTAGIVSAKGRSLPNESTPTFTQDVA 240  
 Db 181 teelpvkvkignpknlgpewvaigapfgpdnsvtagivsakgrslpnesytpftqtdva 240  
 QY 241 INPNSGGPLFNLKGQVVGINSQIYRSRGGFMGISFAIPIDVAMNVAEQKNTGKVORQG 300  
 Db 241 inpgnsggplfnlkgvgvngsqisrsggfmgisfaipidvamnvaekntgkvorgq 300  
 QY 301 LGVIEVSYGLAQSGFLDKASGALIAKILPSPAPERAGLQAGDIVLSLDGGEIRSSGDL 360  
 Db 301 lgvieqvsyglagsgfldkaggaliaakilpgspaperagiqagdivlsldggeirssgdl 360  
 QY 361 PVMVGAITPCKEYSLGWRKGEETIKAKLGNAAEHTGASSKTDEAPYTEOQSGTFESVES 420  
 Db 361 pvmvgaaitpckeyslgwrkgeetikaklgnaaehtgassktdeapYTEOQSGTFESVES 420  
 QY 421 AGTTLQHTDSSGKHLVVRVSDAAERAGLRHGDIEILAVRASP 463  
 Db 421 agttlqhtdssgghlvvrsvsdaaeraglrghdieleilavrgvp 463  
 RESULT 5  
 AAY75750  
 ID AAY75750 standard; Protein; 499 AA.  
 XX  
 AC AAY75750;

XX 21-MAR-2000 (first entry)  
 XX Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2972.  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy.  
 XX Neisseria meningitidis.  
 OS Neisseria meningitidis.  
 XX WO9957280-A2.  
 PN 11-NOV-1999.  
 PD 30-APR-1999; 99WO-US09346.  
 PF 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX WPI: 2000-062150/05.  
 DR N-PSDB; AA254512.  
 XX Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics -  
 PT Claim 2; Page 1392-1393; 1453pp; English.  
 PS  
 XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 XX Sequence 499 AA;  
 SQ  
 Query Match 95.4%; Score 2244; DB 21; Length 499;  
 Best Local Similarity 95.9%; Pred. No. 2.5e-181;  
 Matches 444; Conservative 5; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 VFKKYQYFALAAALCAGCERAGSFFGADKKREASVERIEHTKDDGSVSMILPDPFAQL 60  
 Db 1 vfkkgyalaaalcaaslagckagsgffgadkkaesveriehkddgsvsmilpdtfaql 60  
 QY 61 VQSEGPVAVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMPEIQEEADD 120  
 Db 61 vqsegpavvniqaaprtqngsgnaetdsdpladndpfyefkrlvpmpeiqeeadd 120  
 QY 121 GGLNFGSGFIISKNGYILTNTHVVMGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180  
 Db 121 gglnfgsgfiiskngyiltnthvvtgmgsikvllndkreytakligsdvqsdvallkida 180

QY 181 TEELPVVKIGNPKNLKPCGEVAAIGAPGFDNSVTAGIVSAKGRSLPNESYTPFIOTDVA 240  
Db 181 teelpvkvignpdklpgevwaaigapfgfndsvtagvsaakgrslpnesytpfiqtdva 240  
QY 241 INPNSGGPLNLKGQVGVINSQIYSRSGFMGISFAIPIDVAMNVAEOLKNTGKVORGO 300  
Db 241 inpnsggplfnlkgvvgvlnsqiysrsgfmgisfaipidvamnvaeklntgkvqrgq 300  
QY 301 LGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360  
Db 301 lgviiqevsyglasfgldkasgaliakilpgspaeragragdivlsldggeirssgdl 360  
QY 361 PMVGAITPGKEVSLGVWRKGEETIKAKLGNAAEHTGASSKTDAPYTEQSGTFSVES 420  
Db 361 pvmvgaitpgkevsldgvwrkgeetikaklgnaaehigassktdapYTEQSGTFSVES 420  
QY 421 AGITLQHTDSSGKHLVVVRVSDAERAGRLRHGDEILAVRAS 463  
Db 421 agitlqhtdssgghlvvrsvdaaeragrlrrgdellavagvp 463

RESULT 6

AY70413  
ID AAY70413 standard; Protein; 498 AA.

AC AAY70413;

DT 03-JUL-2000 (first entry)

DE Neisseria meningitidis NMASP protein-2.

KW NMASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
KW anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicaemia.  
OS Neisseria meningitidis.

\*PH Key Location/Qualifiers

FT Misc-difference 8 /note= "Encoded by TTC"

FT Misc-difference 17 /note= "Encoded by TTG"

FT Misc-difference 22 /note= "Encoded by GAA"

FT Misc-difference 42 /note= "Encoded by GAA"

FT Misc-difference 58 /note= "Encoded by GCC"

FT Misc-difference 84 /note= "Encoded by GGC"

FT Misc-difference 134 /note= "Encoded by AAC"

FT Misc-difference 145 /note= "Encoded by GCC"

FT Misc-difference 194 /note= "Encoded by AAT"

FT Misc-difference 218 /note= "Encoded by GGCATC"

FT Misc-difference 321 /note= "Encoded by AGC"

FT Misc-difference 340 /note= "Encoded by AGG"

FT Misc-difference 387 /note= "Encoded by GCC"

FT Misc-difference 396 /note= "Encoded by ACC"

FT Misc-difference 433 /note= "Encoded by AAA"

FT Misc-difference 451 /note= "Encoded by CAC"

FT Misc-difference 459 /note= "Encoded by AGG"

FT Misc-difference 460 /note= "Encoded by GCA"  
FT Misc-difference 461 /note= "Encoded by AGT"  
FT Misc-difference 463 /note= "Encoded by CGT"  
FT Misc-difference 464 /note= "Encoded by CAA"  
FT Misc-difference 465..498 /note= "Encoded by AGGCAAGTCCCGCTCAA"  
XX WO200012535-A2.  
XX 09-MAR-2000.  
XX 01-SEP-1999; 99WO-US19663.  
XX 01-SEP-1998; 98US-0098685.  
XX (ANTE-) ANTEX BIOLOGICS INC.  
XX Jackson WJ, Harris AM;  
XX WPI: 2000-256581/22.  
XX N-PSDB; AAZ51538.  
XX Neisseria meningitidis NMASP polypeptide, nucleotide sequences and  
XX antibodies, useful in vaccines against infection  
XX Claim 5; Page 70-71; 75pp; English.

XX The present sequence is the Neisseria meningitidis NMASP protein.

XX NMASP is a non-cytosolic protein, with antibacterial and

XX antiinflammatory activity. It shows sequence similarity to E. coli

XX DegP (HtrA) protein. NMASP proteins can be used as ligands to detect

XX antibodies elicited in response to N. meningitidis infections. Cytotoxic

XX anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA

XX may be used for diagnosis, therapy or prophylaxis of Neisserial

XX infections such as, bacterial meningitidis and septicaemia.

XX Sequence 498 AA;

XX Query Match 95.0%; Score 2234.5; DB 21; Length 498;

XX Best Local Similarity 95.3%; Pred. No. 1.6e-180;

XX Matches 444; Conservative 5; Mismatches 13; Indels 1; Gaps 1;

QY 1 VFKKYQYFALAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGVSVMLLPDFAQOL 60

Db 1 vfkkyyqyfalaaalcaaslagcdkagsffgadkkeasfverikhtkddgsvsmllpdfvql 60

QY 61 VQSEGPVAVNIOAAPAPRTONGSGNAETSDSDPLADSDPFYEFFKRLVPMPEIQEEADD 120

Db 61 vqsegpavvniqaapartngsgnaetdsdpladsdpfyeffkrlvpmpeiqeeadd 120

QY 121 GGLNFGSGFTIISKNGYILTNTHVVAGMGSIKVLNDRKREYTAKLIGSDVQSDVALLKIDA 180

Db 121 gglnfgsgftiiskngyiltnthvvtgmgsikvlndrkreytakligsdvqsdvallkida 180

QY 181 TEELPVVKIGNPKNLKPGEWVAAIGAPFGFDNSVTAGIVSAKGRSLPNESYTPFIOTDVA 240

Db 181 teelpvkvignpdklpgevwaaigapfgfndsvtag-vsaakgrslpnesytpfiqtdva 239

QY 241 INPNSGGPLFNLKGQVGVINSQIYSRSGFMGISFAIPIDVAMNVAEOLKNTGKVORGO 300

Db 241 inpnsggplfnlkgvvgvlnsqiysrsgfmgisfaipidvamnvaeklntgkvqrgq 299

QY 301 LGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360

Db 301 lgviiqevsyglasfgldkasgaliakilpgspaeragragdivlsldggeirssgdl 359

QY 361 PMVGAITPGKEVSLGVWRKGEETIKAKLGNAAEHTGASSKTDAPYTEQSGTFSVES 420

Db 361 pvmvgaitpgkevsldgvwrkgeetikaklgnaaehigassktdapYTEQSGTFSVES 420

Db 360 pvmvgaitpgkevalgsvrkgseitkvlklnaehigassktddeapyeqsgtfsves 419

QY 421 AGITLOTHTDSSGRHLVVYRVSDAERAGLRHGDILAVRASP 463

Db 420 agitlqthtdssghlvvrvsdaeraglrghdeilavgvqp 462

RESULT 7

ID AAY52994 standard; Protein; 499 AA.

XX AC AAY52994;

XX DT 21-FEB-2000 (first entry)

XX DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.

XX KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;

XX OS antibiotic; upper respiratory tract infection; bacteraemia; meningitis;

XX PN invasive bacterial disease; antibacterial.

XX OS Neisseria meningitidis.

XX PN WO9955872-A1.

XX PD 04-NOV-1999.

XX PF 20-APR-1999; 99WO-EP02765.

XX PR 23-APR-1998; 98GB-0008734.

XX \*A (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Ruelle J;

XX DR WPI: 2000-052809/04.

XX DR N-PSDB; AA233306.

XX PT Novel polynucleotides and polypeptides from Neisseria meningitis used

XX PT to prepare vaccines against bacterial infections -

XX PS Claim 3; Page 77-78; 94pp; English.

XX CC The present sequence represents a BASB013 polypeptide isolated from

XX CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be

XX CC employed as research reagents and material for the discovery of

XX CC treatments and diagnostics for diseases, particularly human diseases.

XX CC They can be used for diagnosis of disease, staging of disease, or

XX CC determining response of an infectious organism to drugs. The

XX CC polynucleotides may be used as a source for hybridisation probes, and

XX CC for screening of genetic mutations, serotype, organism or strain

XX CC identification, identification of mutation in BASB013 sequences, and as

XX CC components of arrays which are useful for diagnostic and prognostic

XX CC purposes. The polypeptides can be used to produce antibodies. The

XX CC polypeptides can also be used in vaccine formulations, and to identify

XX CC agonists and antagonists. The polypeptides, antibodies, agonists and

XX CC antagonists (which are bacteriostatic) are used for the treatment and

XX CC prevention of diseases such as upper respiratory tract infection,

XX CC invasive bacterial diseases such as bacteraemia and meningitis, and for

XX CC the development and screening of antibacterial drugs. They are also used

XX CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins

XX CC on in-dwelling devices, or to extracellular proteins on wounds, and to

XX CC thus prevent tissue damage and/or block the normal progression of

XX CC pathogenesis in infections initiated other than by the implantation of

XX CC in-dwelling devices or by other surgical techniques.

XX SQ Sequence 499 AA;

Query Match 93.6%; Score 2202; DB 21; Length 499;

Best Local Similarity 94.0%; Pred. No. 9.1e-178;

Matches 435; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 VFKKYQYFALAAALCAALLACCEKAGSFFGADKKKEASFEVRIETHTKDDGSVSMLLPDFAQL 60

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

1 mfkkyqylalaaicaaslagckagsffgadkkeastverlehckddgsvsmllpdfvql 60

QY 61 VQSEGPAVVNIQAAPARTQNGSCNAETSDSDPLADSDPFYEFFKRLVPMPEIQEEADD 120

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

61 vqsegpavvniqaapartqngsnaetdsdpladsdpfyeffkrlvpmpeiqeeadd 120

QY 121 GGLNFGSGFIISKNGYILTNTHVVAGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

121 gglnfgsgfiiskngyiltnthvvtgmgsikvllndkreytakligsdvqsdvallkida 180

QY 181 TEELPVVYKIGNPKNLKPGEWAAIGAPFGFDSVNTAGIVSAKGRSLPNESYTPFTQTDVA 240

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

181 teelpvvkignpkdlkpgewaaigapfgfdnsvtagivsaakgrslpnesytpfqtqtdva 240

QY 241 INPNSGGPLFLNLKGOVVGINSQIYSRSGFGMSIFAIPIDVAMNVAEOLKNTCKVORGQ 300

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

241 inpgnsggplfnlkqvgvnginsqlysrsgfgmgsifalpidvamnvaeeqlkntgkvrgq 300

QY 301 LGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

301 lgviiqevsyglasqsgfldkagallakilpgspaeragilqagdivlsldgeirssgdl 360

QY 361 PVMVGAITPGKEVSLGVWRKGEETIKAKLGNAEHTGASKTDEAPYTEQOOSTFSVES 420

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

361 pvmvgaitpgkevsylgvwrkgeetikaklgnaehtgasktdeapyteoostfsves 420

QY 421 AGITLOTHTDSSCKHLVVVVRVSDAERAGLRHGDILAVRASP 463

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

421 agitlothtdssckhlvvvrvsdaeraglrghdeilavrgqp 463

RESULT 8

AAY52993

ID AAY52993 standard; Protein; 499 AA.

XX AC AAY52993;

XX DT 21-FEB-2000 (first entry)

XX DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.

XX KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;

XX OS antibiotic; upper respiratory tract infection; bacteraemia; meningitis;

XX PN invasive bacterial disease; antibacterial.

XX OS Neisseria meningitidis.

XX PN WO9955872-A1.

XX PD 04-NOV-1999.

XX PF 20-APR-1999; 99WO-EP02765.

XX PR 23-APR-1998; 98GB-0008734.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Ruelle J;

XX DR WPI: 2000-052809/04.

XX DR N-PSDB; AA233306.

XX PT Novel polynucleotides and polypeptides from Neisseria meningitis used

XX PT to prepare vaccines against bacterial infections -

XX PS Claim 5; Page 74-76; 94pp; English.

XX CC The present sequence represents a BASB013 polypeptide isolated from

XX CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be

XX CC employed as research reagents and material for the discovery of

XX CC treatments and diagnostics for diseases, particularly human diseases.

CC They can be used for diagnosis of disease, staging of disease, or  
CC determining response of an infectious organism to drugs. The  
CC polynucleotides may be used as a source for hybridisation probes, and  
CC for screening of genetic mutations, serotype, organism or strain  
CC identification, identification of mutation in BASB013 sequences, and as  
CC components of arrays which are useful for diagnostic and prognostic  
CC purposes. The polypeptides can be used to produce antibodies. The  
CC polypeptides can also be used in vaccine formulations, and to identify  
CC agonists and antagonists. The polypeptides, antibodies, agonists and  
CC antagonists (which are bacteriostatic) are used for the treatment and  
CC prevention of diseases such as upper respiratory tract infection,  
CC invasive bacterial diseases such as bacteraemia and meningitis, and for  
CC the development and screening of antibacterial drugs. They are also used  
CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
CC on in-dwelling devices, or to extracellular proteins on wounds, and to  
CC thus prevent tissue damage and/or block the normal progression of  
CC pathogenesis in infections initiated other than by the implantation of  
CC in-dwelling devices or by other surgical techniques.

XX Sequence 499 AA;

Query Match 93.4%; Score 2197; DB 21; Length 499;  
Best Local Similarity 93.7%; Pred. No. 2.4e-177;  
Matches 434; Conservative 10; Mismatches 19; Indels 0; Gaps 0;

Qy 1 VFKKYQYFALAALCAALCKAGSFFGADKKEASFVERIEHTKDDGSVSMLLPDPFQAL 60  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
1 mfkkyglalaaalcaaslagcdkagsfgadkkeasfveriehtkddgsvsmllpdpfvl 60  
Qy 61 VQSEGPVAVNIQAAPRTQNSGNAETSDPLADSDPFYEFKRLVPMPEIPOEADD 120  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
61 vqsegpavvniqaaprtqnsгнаetdspladsdpfyefkrlvpmmpaipgeadd 120  
Qy 121 GGLNFGSGFIISKNGYILNTHVWAGMSIKVLLNDKREYAKLIGSDVSDVALLKIDA 180  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
121 gglngsgfllskdgyllnthvwtgmsikvllndkreytaklgsdvsdvallkida 180  
Qy 181 TTELPVVKIGNPKNLKPGEWAAIGAAPPFGFNSVTAGIVSAKGRSLPNESYTFPIQTDVA 240  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
181 teelpvkvignpkdkpgewaaigprfgfndsvtagivsaksrslpnesytpfiqtdva 240  
Qy 241 INPGNSGGPLNLKGOVGINQIVSRSGFGMGISFAIPIDVAMNVAEQLNKTKVORGQ 300  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
241 inpgnsggplnlkgvgvngsqivrsrgfmgisfaipidvammvaeqlnktgkvrgq 300  
Qy 301 LGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
301 lgvitqevsyglaqsfldkaggaliaakilpgspaeraglaqgdvllslldggeirssgd 360  
Qy 361 PVMWGAITPGKEVSLGWWRKGBEITTKAKLGNAAEHTGASSKTDAPYEQQSGTFSVES 420  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
361 pvmwgaitypgkevsigwvrkgeitvavklgnaseqtgsssepdkypaehqsgtfsves 420  
Qy 421 AGITLQTHFTDSGKHLVVVRVSDAERAGLRHGHDEILAVRAS 463  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||  
421 agitlqthdsggrlvvrsvgaaeraglrghgdeillavrgvp 463

RESULT 9  
AAV70414  
ID AAV70414 standard; Protein; 475 AA.  
XX  
AC AAV70414;  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE Neisseria meningitidis NMASP protein-3.  
XX  
KW NMASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
KW anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicaemia.  
XX

OS Neisseria meningitidis.  
XX Key Location/Qualifiers  
FH Misc-difference 19 /note= "Encoded by GAA"  
FT Misc-difference 35 /note= "Encoded by GAA"  
FT Misc-difference 61 /note= "Encoded by GGC"  
FT Misc-difference 111 /note= "Encoded by GGC"  
FT Misc-difference 122 /note= "Encoded by AAC"  
FT Misc-difference 171 /note= "Encoded by GGC"  
FT Misc-difference 194 /note= "Encoded by AAT"  
FT Misc-difference 298 /note= "Encoded by GGCATC"  
FT Misc-difference 317 /note= "Encoded by AGC"  
FT Misc-difference 364 /note= "Encoded by AGG"  
FT Misc-difference 373 /note= "Encoded by GGC"  
FT Misc-difference 410 /note= "Encoded by ACC"  
FT Misc-difference 428 /note= "Encoded by AAA"  
FT Misc-difference 436 /note= "Encoded by CAC"  
FT Misc-difference 437 /note= "Encoded by AGG"  
FT Misc-difference 438 /note= "Encoded by GCA"  
FT Misc-difference 440..475 /note= "Encoded by AGT"  
XX /note= "Encoded by CGTCAA"  
WO200012535-A2.  
XX  
XX 09-MAR-2000.  
XX 01-SEP-1999; 99WO-US19663.  
XX 01-SEP-1998; 98US-0098685.  
XX (ANTE-) ANTEX BIOLOGICS INC.  
XX Jackson WJ, Harris AM;  
XX WPI; 2000-256581/22.  
XX N-PSDB; AAZ51539.  
XX Neisseria meningitidis NMASP polypeptide, nucleotide sequences and  
XX antibodies, useful in vaccines against infection -  
XX Claim 5; Page 71-73; 75pp; English.  
XX  
XX The present sequence is the Neisseria meningitidis NMASP protein.  
XX NMASP is a non-cytosolic protein, with antibacterial and  
XX antiinflammatory activity. It shows sequence similarity to E. coli  
XX DegP (HtrA) protein. NMASP proteins can be used as ligands to detect  
XX antibodies elicited in response to N. meningitidis infections. Cytotoxic  
XX anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA  
XX may be used for diagnosis, therapy or prophylaxis of Neisserial  
XX infections such as, bacterial meningitis and septicaemia.  
SQ Sequence 475 AA;

Query Match 90.5%; Score 2130.5; DB 21; Length 475;  
Best Local Similarity 96.4%; Pred. No. 9.6e-172;  
Matches 424; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

QY 24 AGSFFGADKKEASFVERIEHTKDDGSVMLLPDPFAQLVQSEGPVAVNIQAAPAPRTQMS 83  
DB 1 agsffgadkkeasfveriehtkddgsvmllpdpfvqlvqsegpavvniqaapaprtqms 60  
QY 84 GNAETSDPLADSDPYEFKRLVPNMPEIQBEADGGLNFGSGFTISKNGYILTNTHV 143  
DB 61 snaetdspladsdpyeffkrlvpnmpeipqeeaddggnfsgfllskdgyllntnvh 120  
QY 144 VAGMGKIVLLNDKREYAKLIGSDVQSVALLKIDATEELPVVKIGNPKLPGEWVAA 203  
DB 121 vtmgskivllndkreytakligsdvqsdvallkideateelpvkvignpkdkgewvaa 180  
QY 204 IGAPFGFDSVTVAGIYSAKGRSLPNSYTPFIQTDVAINFNSGGPLFNLKGQVVGINSQ 263  
DB 181 igapfgfdsvtgag-ysakgrslpnsytpfiqtdvaainpgnsggplfnlkgqvvginsq 239  
QY 264 IYSRSGFGWGISFAIPIDVAMNVAEQLNKTKVQSGVQVGGGLVLIQVSYGLAQSGFLDKASG 323  
DB 240 IYSRSGFGWGISfaipidvamnvaeqlntkgvqrgqlvliqevsyglaqsgfldkag 299  
QY 324 ALIAKILPGSPARAGLQAGDIVLSLDGGGIRSGDLPVMVGAIPTGKEYSGLVWRKGE 383  
DB 300 aliakilpgspaeragragdivlsldgggirssgdipvmvgaitpgkeyslgvwrkgee 359  
QY 384 ITTAKLGNAAHTGASSTDEAPYTEQSGTFSVESAGITLQHTDSSGKHLVVRVSD 443  
DB 360 ittkvlgnaaehigassktdeapyleqsgtfsvesagitlqhtdssgghlvvrvsd 419  
QY 444 AAERAGLRHGDDELAVRAS 463  
DB 420 aaeraglrhgdellavragvqp 439

## RESULT 10

AA83151  
ID AAY83151 standard; Protein; 414 AA.

XX  
AC AAY83151;

XX  
DT 24-JUL-2000 (first entry)

XX  
DE NGSP polypeptide of Neisseria gonorrhoeae.

XX  
KW NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
cellular matrix; adherence domain; ligand; detection; diagnosis;  
screening; probe; primer; prophylaxis; therapy.

XX  
PN Neisseria gonorrhoeae.

XX  
PD WO200012133-A1.

XX  
PE 09-MAR-2000.

XX  
PF 01-SEP-1999; 99WO-US20070.

XX  
PR 01-SEP-1998; 98US-0098685.

XX  
PA (ANTE-) ANTEX BIOLOGICS INC.

XX  
PI Jackson WJ, Harris AM;

XX  
DR WPI; 2000-237782/20.

XX  
DR N-PSDB; AA293415.

XX  
PT Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
Neisseria useful for diagnosis, prevention or treatment of Neisseria  
infections

XX  
PS Claim 5; Page 63-64; 68pp; English.

XX  
CC The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
Arg-Gly-Asp motifs near the C-terminus which function as adherence

CC domains for extracellular matrix proteins. Using the NGSP polypeptide  
as a vaccine produces antibodies which inhibit binding of N.  
gonorrhoeae to the host's cellular matrix reducing attachment and/or  
subsequent invasion. The NGSP polypeptide and its peptide fragments  
can be used to immunise an animal and produce an immune response.  
They can also be used as ligands to detect antibodies elicited in  
response to Neisseria infections and also as antigens or immunogens  
for inducing Neisseria-specific antibodies which are useful in  
immunassays to detect Neisseria in biological specimens. Nucleotides  
encoding NGSP or its fragments can be used as probes to identify  
Neisseria in biological specimens by hybridization or polymerase  
chain reaction amplification. The NGSP polypeptide can also be used  
in screening assays to identify agents and compounds which useful as  
diagnostic, prophylactic or therapeutic agents against Neisseria  
infection.

XX  
SQ Sequence 414 AA;

Query Match 88.9%; Score 2092; DB 21; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.4e-168;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 MLLPDPFAQLVQSEGPVAVNIQAAPAPRTQSGNAETSDPLADSDPYEFKRLVPNM 111  
DB 1 mllpdpfaqlvqsegpavvniqaapaprtqsgnaetdspladsdpyeffkrlvpnm 60  
QY 112 EIPOEADDGGLNFGSGFTISKNGYILTNTHVAGMGKIVLLNDKREYAKLIGSDVQ 171  
DB 61 elpoeaddggnfsgfllskngyiltnthvagmgskivllndkreytakligsdvq 120  
QY 172 DVALLKIDATEELPVVKIGNPKLPGEWVAAIGAPFGFDSVTVAGIYSAKGRSLPNE 231  
DB 121 dvallkideateelpvkvignpkpgewvaaigapfgfdsvtgivaakgrslpnes 180  
QY 232 TPFITQDVAINFNSGGPLFNLKGQVVGINSQIYSGFGWGISFAIPIDVAMNVAEQ 291  
DB 181 tpfitqdvainpgnsggplfnlkgqvvginsqliysrsgfmgisfaipidvamnvae 240  
QY 292 NTGKVQVQGLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLS 351  
DB 241 ntgkvqrgqlgvliqevsyglaqsgfldkasgalliakilpgspaeraglaqdiv 300  
QY 352 GEIRSGDLPVMVGAIPTGKEYSGLVWRKGEITTKAKLGNAAHTGASSTDEAPYTE 411  
DB 301 geirssgdipvmvgaitpgkeyslgvwrkgeetikaklignaahtgassktdeap 360  
QY 412 QSGTFSVESAGITLQHTDSSGKHLVVRVSDAAERAGLRHGDDELAVRASPRQ 465  
DB 361 qsgtfsvesagitlqhtdssgkhlvvrvsdaaeraglrhgdellavrasprq 414

## RESULT 11

AA870409

ID AAY70409 standard; Protein; 448 AA.

XX  
AC AAY70409;

XX  
DT 03-JUL-2000 (first entry)

XX  
DE Neisseria meningitidis NMASP protein-1.

XX  
KW NMASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
Neisserial infection; meningitidis; septicaemia.

XX  
OS Neisseria meningitidis.

XX  
FH Key . Location/Qualifiers

FT Misc-difference 167

XX  
TX /note= "Encoded by NTC"

XX  
PN WO200012535-A2.



XX 09-MAR-2000.  
PD 01-SEP-1999; 99WO-US19663.  
XX 01-SEP-1998; 98US-0098685.  
XX (ANTE-) ANTEX BIOLOGICS INC.  
XX Jackson WJ, Harris AM;  
PI WPI: 2000-256581/22.  
XX N-PSDB: AAZ51533.  
XX  
PT Neisseria meningitidis NMASP polypeptide, nucleotide sequences and  
PT antibodies, useful in vaccines against infection -  
XX  
PS Claim 5; Page-; 75pp; English.  
XX  
CC The present sequence is the Neisseria meningitidis NMASP protein.  
CC NMASP is a non-cytosolic protein, with antibacterial and  
CC anti-inflammatory activity. It shows sequence similarity to E. coli  
CC DegP (HtrA) protein. NMASP proteins can be used as ligands to detect  
CC antibodies elicited in response to N. meningitidis infections. Cytotoxic  
CC anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA  
CC may be used for diagnosis, therapy or prophylaxis of Neisserial  
CC infections such as, bacterial meningitis and septicemia.  
CC Note: The protein represented in SEQ ID NO:2 of the specification is  
CC erroneous. The present sequence is the decoded version of the nucleotide  
CC represented in AAZ51533.  
XX  
SQ Sequence 448 AA;  
  
Query Match 85.1%; Score 2002; DB 21; Length 448;  
Best Local Similarity 96.4%; Pred. No. 6.6e-161;  
Matches 397; Conservative 3; Mismatches 12; Indels 0; Gaps 0;  
  
Qy 52 MLLPFAQLVQSEGPAVNVNIAAPRTONGSGNAETSDPLADSPFFVEFFKRLVNPMP 111  
Db 1 mlppdvqlvqsegpavvniqaaprtngssnaetdsdpladsdpffveffkrlvnpmp 60  
  
Qy 112 EIPOEADDGGLNFCSGFTISKNGYILTNTHVVAGMGSIKVLLNDKREYAKLIGSDVQS 171  
Db 61 elpoeaaddgglngsfgtiskngyiltnthvvtgmsikvllndkreytakligsdvqs 120  
  
Qy 172 DVALLKIDATEELPVYKIGNPNKLPGEWVAAGAPFGFEDNSVTAGIVSAKGRSLPNESY 231  
Db 121 dvallkideateelpvkkignpkdlkpgewvaagapfgfdnsvtagvsakgrslpnesy 180  
  
Qy 232 TPFTQTDVAINPGNSGGPLFNLKGQVVGINSQIYRSRGGFMGLSFAIPIDVAMNVAEQLK 291  
Db 181 tpfqtvdvaipngnsggplfnlkgqvvginsqiyrsrggfmglisfaipidvamnvaekl 240  
  
Qy 292 NTGKVQRGOLGVIQEVSVGLAQSGELDKASCALITAKILPGSPAERAGLQAGDIVLSLDG 351  
Db 241 ntgkvqrqglvileqsvyglqsgelgdkagallakilpgspaeragragdivlsldg 300  
  
Qy 352 GEIRSSGDLPMVWGAIITPGKEVSLGVWRKGEETITKAKLGNAAEHTGASSKTEAPYTEQ 411  
Db 301 geirssgdlpmvvgaitpgkevsylvwrkgeetilkvkglnaaehlgasskdeapYTEQ 360  
  
Qy 412 QSGTFSVESAGITLTQHTDSSGKHLVWVRVSDAAERAGLRHGDIELAVRASP 463  
Db 361 qsgtfsvesagitltqhtdssghlvvvrsvdaaeraglrghgdellavvgvp 412  
  
RESULT 12  
AA52996  
ID AA52996 standard; Protein: 370 AA.  
XX  
AC AA52996;  
XX

DT '21-FEB-2000 (first entry)  
XX  
DE Neisseria meningitidis BASB013-C protein sequence.  
XX  
KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9955872-A1.  
XX  
PD 04-NOV-1999.  
XX  
PF 20-APR-1999; 99WO-EP02765.  
XX  
PR 23-APR-1998; 98GB-0008734.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Ruelle J;  
XX  
DR WPI: 2000-052809/04.  
DR N-PSDB: AAZ33308.  
XX  
PT Novel polynucleotides and polypeptides from Neisseria meningitis used  
PT to prepare vaccines against bacterial infections -  
XX  
PS Example 2; Page 82-83; 94pp; English.  
XX  
CC The present sequence is a conserved BASB013-C polypeptide isolated  
CC from Neisseria meningitidis. BASB013 polynucleotides and polypeptides  
CC may be employed as research reagents and material for the discovery of  
CC treatments and diagnostics for diseases, particularly human diseases.  
CC They can be used for diagnosis of disease, staging of disease, or  
CC determining response of an infectious organism to drugs. The  
CC polynucleotides may be used as a source for hybridisation probes, and  
CC for screening of genetic mutations, serotype, organism or strain  
CC identification, identification of mutation in BASB013 sequences, and as  
CC components of arrays which are useful for diagnostic and prognostic  
CC purposes. The polypeptides can be used to produce antibodies. The  
CC polypeptides can also be used in vaccine formulations, and to identify  
CC agonists and antagonists. The polypeptides, antibodies, agonists and  
CC antagonists (which are bacteriostatic) are used for the treatment and  
CC prevention of diseases such as upper respiratory tract infection,  
CC invasive bacterial diseases such as bacteraemia and meningitis, and for  
CC the development and screening of antibacterial drugs. They are also used  
CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
CC on in-dwelling devices, or to extracellular proteins on wounds, and to  
CC thus prevent tissue damage and/or block the normal progression of  
CC pathogenesis in infections initiated other than by the implantation of  
CC in-dwelling devices or by other surgical techniques.  
XX  
SQ Sequence 370 AA;  
  
Query Match 77.9%; Score 1832; DB 21; Length 370;  
Best Local Similarity 97.0%; Pred. No. 1.2e-146;  
Matches 359; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 1 VFKKYQYFALAAALACGACEKAGSFFGADKKKEASFVERIETHTKDDGVSMLLPFAQL 60  
Db 1 mfkkyqylalaaalcaaslagcdkagsffgadkkeasfveriehtkddgvsmlldpfaql 60  
  
Qy 61 VQSGPAAVNTQAAPRTONGSGNAETSDPLADSPFFVEFFKRLVNPMPPEAEADD 120  
Db 61 vqsgpavvniqaaprtngsgnaetdsdpfadspffveffkrlvnpmppeaeadd 120  
  
Qy 121 GGLNFGSGFTISKNGYILTNTHVVAGMGSIKVLLNDKREYAKLIGSDVQSVALLKIDA 180  
Db 121 gglngsgftiskngyiltnthvvtgmsikvllndkreytakligsdvqsvallkida 180  
  
Qy 181 TEELPVYKIGNPNKLPGEWVAAGAPFGFEDNSVTAGIVSAKGRSLPNESYTPFTQTDVA 240





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:21 ; Search time 16.36 Seconds  
(without alignments)  
694,248 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQYFALALCAALLAG.....ERAGLRHGEITLAVRASPRQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	990.5	42.1	460	4	US-09-199-637A-132
2	731.5	31.1	503	1	US-08-245-294-8
3	731.5	31.1	503	1	US-08-474-499-8
4	731.5	31.1	503	1	US-08-307-279A-8
5	731.5	31.1	503	5	PCT-US95-06211-8
6	646	27.5	463	1	US-08-485-569-2
7	646	27.5	463	1	US-08-480-993-2
8	646	27.5	463	2	US-07-903-079B-2
9	633.5	26.9	475	1	US-08-350-741-2
10	633.5	26.9	475	2	US-08-463-875A-2
11	632	26.9	463	1	US-08-278-091-2
12	632	26.9	463	1	US-08-483-859-2
13	632	26.9	463	1	US-08-472-173-2
14	632	26.9	463	2	US-08-487-167-2
15	632	26.9	463	2	US-08-482-816-2
16	632	26.9	463	2	US-08-296-149-2
17	632	26.9	463	2	US-08-801-499-2
18	632	26.9	463	2	US-08-615-271-2
19	632	26.9	463	3	US-09-074-660-2
20	632	26.9	463	3	US-09-074-659-2
21	632	26.9	463	3	US-09-106-468-2
22	632	26.9	463	4	US-09-106-466A-2
23	632	26.9	463	4	US-09-106-467-2
24	631.5	26.8	475	1	US-08-278-091-6
25	631.5	26.8	475	1	US-08-483-859-6
26	631.5	26.8	475	1	US-08-472-173-6
27	631.5	26.8	475	2	US-08-487-167-6

ALIGNMENTS

RESULT 1  
US-09-199-637A-132  
; Sequence 132, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hul  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 132  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-132

Query Match	42.1%	Score	990.5	DB	4	Length	460
Best Local Similarity	52.5%	Pred. No.	2.6e-84				
Matches	220	Conservative	55	Mismatches	111	Indels	33
Gaps	11						
Qy	54	LPDFAQLVQSEGPAYVNI	---QAAPAPRTQNGSGNAETSDPLADSDP	-FYEFKRLVFN	109		
Db	27	LPDFTPLVEQASPAVVNISTROKLP	-----DRAMARQLSIPDLEGLPPMFRDLERTIPQ	82			
Qy	110	MPEIP	---QEADGGLNFGSGFIISKNYI	LTNTHVAGMSIKVLLNDKREYTA	166		
Db	83	VPRNPGQREAQ	-----SLGSGFIISNDGYI	LTNNHVADAEILVRLSDRSEHKAKLVG	138		
Qy	167	SDVQSDVALLKIDATEELPVVYKGNPKNLKPGEWAAICAPFGFNSVTAGIVSAKGRSL	226				
Db	139	ADPRSDVAVLKIEA	-KNLPTKLGSNKLKVGWEVLAIGSPFGFDSHTVAGIVSAKGRSL	197			
Qy	227	PNESYTFPIQTDVAINPONGSGPLNLRKGVVGINSQIYSRSGGFMGISFAIPIDVAMNY	286				
Db	198	PNESYVFPFIQTDVAINPONGSGPLNLRKGVVGINSQIYSRSGGFMGISFAIPIDVAMNY	257				
Qy	287	AEQLKNTGKVGQGLVYIIQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIV	346				

```
Db 258 ADQLKAGKVSRLVGLVGVVIOEVNKNDAESFGLDRKPSGALVAQLVEDGPAARKGLQVGDI 317
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 347 LSLDGGIRSSGDLPMVVGAIPTGKEYSLGVWRKGEETIKAKLGNAAEHTGASSKTD- 405
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 318 LSLNGOSINESADLPHLVGNKPGDKINLVIRNGORKSLSMVGNLPD-----DDEE 370
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 406 -----APYTQQSGTFSVESAGITLQTH--TDSSGKHLVVVRVSDA-AERAGLRHGD 456
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 371 IASMGAPGAERSSNRLGVTVADLTAEORKSLDIQG-GVVIKEVQDGPAAVIGLRPGDI 428
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :

RESULT 2
US-08-245-294-8
; Sequence 8, Application US/08245294
; Patent No. 5644047
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DIAGNOSING
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA
; TITLE OF INVENTION: INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,294
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.612
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-245-294-8

Query Match 31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 5.2e-60;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQLVQSEGA 67
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 15 FSAALETALFFSGC---GSSLWTTKAHANSV-----FSSLMQOQGFADIVSQVKPA 62
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 68 VVNIQAAPRTQN----GSGNAETSDPLADSDP----FYEFFKRLVPNMPEIPQEEADD 120
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 63 VVSQVQKSNKKKEWFFSDFSTPGEDLPDQHPKLFKFFQDFYNRDRKPSNKS L-QRSRL 121
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 121 GGLNFGSGFIISKNGYILNTHVVGMSGSKVLLNDKREYTAKLIGSDVQSDVALLIDA 180
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 122 RPIAFSGGFIISDGGYIVTNHNVISDGTSYAVVLDDGTETLNAKLIGDPRDLAVLKVNE 181
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 181 TEELPVVIGNPKNLKPGEWAAIGAPFGDGNVTAIGVSAKGRSLPNSYTPPIQTDA 240
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
```

```
Db 182 KRKFYSVDFGDDSKLRVGDVWVAIGNPFLGGTGTAGIVSARGRDIGTGVYDDFIQIDAA 241
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 241 INPGSGGPLENLKGOVVGINSOISYRSGGFMGISFAIPDIDVAMNVAEOLKNTCKVORG 300
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 242 VNRSGSGGTFDLNGRVGVNTAIFSPSGNVGIAFAIPAANTANEVVQQLIEKGLVORG 301
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 301 LGVITQEVSYGLAQSFGLDKASGALIAKTILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 302 LGVQIQPVTKESIDSLGKEAKGALLITDPLKG-PAKAGIKAGDVIISVNGEKINDYRDL 360
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 361 PYMVGAITPGKEVSLGVWRKGEETIKAKLGNAAEHTGASSKTDDEAPYT-BOOSGTFSVE 419
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 361 AKRIANMSPGETVTLGVWRKSGKEENIKVLDSPED---ENMKDGSYSNEHGNSETLE 417
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
QY 420 SAGITLQTHTDSSGKHLVVVRV---SDAERAGLRHGDIELAV 459
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 418 DYGLIVAPSDDGLG--LVYTDVDPDSDAADK-GIRPGDVIVTV 457
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :

RESULT 3
US-08-474-499-8
; Sequence 8, Application US/08474499
; Patent No. 5693776
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE
; TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,499
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,294
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.612
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-499-8

Query Match 31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 5.2e-60;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQLVQSEGA 67
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 15 FSAALETALFFSGC---GSSLWTTKAHANSV-----FSSLMQOQGFADIVSQVKPA 62
; :||| ||| ||| ||| ||| : ||| ||| ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
```



```

; TYPE: amino acid
; TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US95-06211-8

```

Query Match	31.1%;	Score 731.5;	DB 5;	Length 503;
Best Local Similarity	39.5%;	Pred. No. 5.2e-60;		
Matches 183;	Conservative	76;	Mismatches 173;	Indels 31;
				Gaps 11;

QY	8	FALAAICALLACEKAGSGFFGADKKEASFEVERIEHTKDDGVSMILLPDPFAQLVQSEGRA	67
Db	15	FSAALETALFFSGC---GSSLWTTKAHANSV-----FSSLMOQOGFADIVSOVKPA	62
QY	68	VVNIOAAPARTON---GSGNAETSDSDPLADSDP---FVEEFKRLVPNNKPEIOEEADD	120
Db	63	VVSVOVKSNNKKKEWFSEDFSTPGFDQLPDQHPKLKFFQDFINRDKPSNKSU-QRSHKL	121
QY	121	GGUNFGSGFTISKNGXYILNTHVYVAGMSGIKVLLNDNKREYTKATLIGSDVQSDVALLKIDA	180
QY	122	RPIAFSGGFFISSDGYIVTNHNVISDGTSVAVVDDPGTELNAKLIGTDPRTDLAVLKVNE	181
QY	181	TEELPVVKTGNPKNKPGEVVAIGAAPGFCFDSNVTAGIYSAGRSPLNPSYTFPTQTDVA	240
Db	182	KRFSTVDFGDDSKLRVGWVVAIGNPFGLGGTVTAGIYSAGKRDIGTGVYDDFIQDA	241
QY	241	INPFGSGGFLNLKGOWGINSQIYSRSGGFMGISPAIPIDVAMNVAEOLKNTKGVQRC	300
Db	242	VNRGSGGPTDLNGKVVGVNTAIPSPSGNVGIAIPAIPAATANEVVQQLIEKGLVQRC	301
QY	301	LGVIIQEVSYGLAQSGLDKASGALIAKILPQSPABRACGLQAGDVISSLDGGEIRSSGDL	360
Db	302	LGVIQIPVTKETSISGLKEARGALITDPLKG--PAAKAGIKAGDVIISVNGEKINDVRDL	360
QY	361	PVMVGAITPCKEYSLGVRKRGEEITIKALGNAAEHTGASSKTDRAPTY-EQSGTFSEVE	419
Db	361	AKRIANNSGCVTVTLGVWKSARGEENIKVKLDSMPED---ENKKDGSKYSNEHGNSDTEIL	417
QY	420	SAGITLQTHDSSGKHLVVYRV---SDAAERAGLRHGDEILAV	459
Db	418	DYGLIIVAPSDDGIG--LVYTDVDPDSDAADK-GIRPGDVIVTV	457

RESULT 6  
US-08-485-569-2  
Sequence 2, Application US/08485569  
Patent No. 5679547  
GENERAL INFORMATION:  
APPLICANT: Krivan, Howard C.  
APPLICANT: Samuel, James E.  
TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE  
TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennile & Edmonds  
STREET: 1155 Avenue of Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485.569  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/903,079  
FILING DATE: 22-JUN-1992  
PRIOR APPLICATION DATA:

```

: APPLICATION NUMBER: US 07/810,966
: FILING DATE: 20-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/631,698
: FILING DATE: 21-DEC-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Baldwin, Geraldine F.
: REGISTRATION NUMBER: 31,232
: REFERENCE/DOCKET NUMBER: 7969-049
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 463 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-485-569-2

```

Query Match	27.58;	Score 646;	DB 1;	Length 463;
Best Local Similarity	38.9%;	Pred. No. 4.5e-52;		
Matches 162;	Conservative 75;	Mismatches 151;	Indels 28;	Gaps 12;
Qy	54	LPDFAQLVQSGPAAVNIQAAPRTONGSGNAETDS--DPLADSDP--FYEFF-KRLVPN	109	
Db	28	LPSEFVSQNSLAPMLEKVQ--PAVVTLVSEGKAKVDSRSPFLDDIPIPEKFKFFGCDRFAEQ	85	
Qy	110	MPEIQBEADDGGLVCGSGFTII-SKNGYILNTHTHVWAGMSGIKVLLNDKREYTKATGLGSD	168	
Db	86	FGRGESKRNPRL--GSGVLIINASKGYLVLTNNHVIDGADKITVQLQDGRFEKAKLVCKD	143	
Qy	169	VOSDVALLKIDATEELPVKIGCNPKNLKPGEBWAAACAPGCFDNSVYAGIYSAKGRSLPN	228	
Db	144	EQSDIALVQLKESNLTEIRFADSDKRVGDTVAIGNPFGLOGTIVTSGIYVSLAGRSTGS	203	
Qy	229	ES--YTFPIQTDVAINPCNSGGLPLNFKGVGVNGSIYSRSGGFMGTSAPIPDVAMNV	286	
Db	204	DSGTENVYIOTDAAVNRGNSGALVNLGELIGINTAIIPSGCGNAGIATAIPSNQASNL	263	
Qy	287	AEQLKNTGKVQRGGLYIIQEVSVYGLAQSGFLDKASGALIAKIIILPGSPAERAGLQAGDIV	346	
Db	264	VQQLTEFGQVRRGLLTKGGEINLADLAKAFNVSAQQAQFAVSEVLPKSAAEKAGLUKAGDII	323	
Qy	347	LSLDGGEIRSGDLPVVMVGATTPCKEYVSLGVWVRKGEETITIKAKLGNAAEHGTSASKYDEA	406	
Db	324	TAMNGQKISSFAETRAKITATTGACKELSLTYLRDQKSHDVKMKL-QADDGSOLSSKT-EL	381	
Qy	407	PYTEQQSGTFSVESAGITTLQHTDSSGKHLVVVRV---SDAAERAGLRHGDIELAV	459	
Db	382	PALD-----GATLKDYDAKGVKGTEITKIQPNLSAAOQ-GLKSGDIITGI	425	

```

RESULT      7
US-08-480-993-2
; Sequence 2, Application US/08480993
; Patent No. 5721115
; GENERAL INFORMATION:
; APPLICANT: Krivan, Howard C.
; APPLICANT: Samuel, James E.
; APPLICANT: No. 5721115berg, Nils T.
; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
;

```



ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480.993  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/903,079  
FILING DATE: 22-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,966  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/631,698  
FILING DATE: 21-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7969-050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-480-993-2

Query Match 27.5%; Score 646; DB 1; Length 463;  
Best Local Similarity 38.9%; Pred. No. 4.5e-52;  
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;  
QY 54 LPDFAQLVQSGPAPVNIQAAPRTONGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109  
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSPFLDDIPEEFKFFGDRFAEQ 85  
QY 110 MPEIQEADGGLNFGSGFII-SKNGYILTNTHVVGMSIKVLLNDKREYTKLIGSD 168  
DB 86 FGGREGSRNPRGL--GSGVIINASKGYVLTNNHVIDGADKITVQLQDGRFKAKLVGKD 143  
QY 169 VQSDVALLKIDATELPVVKIGNPNKLPGEWVAIGAIFGFDNSVTAGIVSAKGRSLPN 228  
DB 144 EQSDIALVQLEKPSNLTEIKFADSKLRVGDFTVAIGNPFLGQTVTSGIVSALGRSTGS 203  
QY 229 ES--YTFPIQTDVAINPGNSGGLNKGQVVGINSQIYSRSGGFMGISFAIPIDVAMNV 286  
DB 204 DSGTYENIQTDAAVNRGNSGALVNLGELIGINTAILSPSGNGAGIAFAPISNOASNL 263  
QY 287 AEQLKNTGKVGQGLGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346  
DB 264 VQIILEFGQVRRLGILKGGELNADLAKAFNVSAQOGAFVSEVLPKSAAEKAGLKAGDII 323  
QY 347 LSLDGGETRSGDLPVMVGAITPGKEVSLGVWRKGEETIKAKLGNAAEHTGASSKTDEA 406  
DB 324 TAMNGOKISSAEIRAKIATTGAGREISLTYLRDCKSHDVKMKL-QADDGSQLSSKT-EL 381  
QY 407 PYTEQSGTFSVESAGITLQTHDSSGKHLVVVRV--SDAAERAGLRHGDITLAV 459  
DB 382 PALD-----GATLKDYDAKGVKGIEITKIQPNSLAAQR-GLKSGDIIIGI 425

RESULT 8  
US-09-388-090-4  
; Sequence 2, Application US/07903079B

Patent No. 5843463  
GENERAL INFORMATION:  
APPLICANT: Krivan, Howard C.  
APPLICANT: Samuel, James E.  
APPLICANT: No 5843463berg, Nils T.  
TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE  
TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/903,079B  
FILING DATE: 22-JUN-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,966  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/631,698  
FILING DATE: 21-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7969-004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-07-903-079B-2

Query Match 27.5%; Score 646; DB 2; Length 463;  
Best Local Similarity 38.9%; Pred. No. 4.5e-52;  
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;  
QY 54 LPDFAQLVQSGPAPVNIQAAPRTONGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109  
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSPFLDDIPEEFKFFGDRFAEQ 85  
QY 110 MPEIQEADGGLNFGSGFII-SKNGYILTNTHVVGMSIKVLLNDKREYTKLIGSD 168  
DB 86 FGGREGSRNPRGL--GSGVIINASKGYVLTNNHVIDGADKITVQLQDGRFKAKLVGKD 143  
QY 169 VQSDVALLKIDATELPVVKIGNPNKLPGEWVAIGAIFGFDNSVTAGIVSAKGRSLPN 228  
DB 144 EQSDIALVQLEKPSNLTEIKFADSKLRVGDFTVAIGNPFLGQTVTSGIVSALGRSTGS 203  
QY 229 ES--YTFPIQTDVAINPGNSGGLNKGQVVGINSQIYSRSGGFMGISFAIPIDVAMNV 286  
DB 204 DSGTYENIQTDAAVNRGNSGALVNLGELIGINTAILSPSGNGAGIAFAPISNOASNL 263  
QY 287 AEQLKNTGKVGQGLGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346  
DB 264 VQIILEFGQVRRLGILKGGELNADLAKAFNVSAQOGAFVSEVLPKSAAEKAGLKAGDII 323  
QY 347 LSLDGGETRSGDLPVMVGAITPGKEVSLGVWRKGEETIKAKLGNAAEHTGASSKTDEA 406

Db 324 TAMNQKISSFAERAKIATTGACKETSILYLRDCKSHDKML-QADDSQSLSKT-EL 381  
QY 407 PYTEQSGTTSVESAGITLTOTDSSGKHLVVVRV---SDAAERAGLRHGDIELAV 459  
Db 382 PALD-----GATLKDYDAKGVKGIEITKIOPNSLAAQR-GLASGDIIIGI 425

## RESULT 9

US-08-350-741-2  
; Sequence 2, Application US/08350741  
; Patent No. 5804194  
; GENERAL INFORMATION:  
; APPLICANT: DOUGAN G.,  
; APPLICANT: CHARLES I.G.,  
; APPLICANT: HORMACHE C.E.,  
; APPLICANT: JOHNSON K.S.,  
; APPLICANT: CHATFIELD S.N.  
; TITLE OF INVENTION: LIVE VACCINES  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON and VANDERHVE PC  
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/350,741  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/952,737  
; FILING DATE: 09-MAY-1994  
; APPLICATION NUMBER: US 07/952,737  
; FILING DATE: 30-NOV-1992  
; APPLICATION NUMBER: GB 9007194.5  
; FILING DATE: 30-MAR-1990  
; APPLICATION NUMBER: PCT/GB91/00484  
; FILING DATE: 28-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILSON, MARY J.  
; REGISTRATION NUMBER: 32,955  
; REFERENCE/DOCKET NUMBER: 117-158  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 475 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-350-741-2

Query Match 26.9%; Score 633.5; DB 1; Length 475;  
Best Local Similarity 34.1%; Pred. No. 7e-51;  
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;  
QY 44 TKDGSVSMILPPFAQLVQSEGPAVV--NQAAAPAPRTQSGNAE----TDSPLADSD 97  
Db 29 TSSAMTAQMPSLAPMLEKVMFSVINVEGSTVTVPMPRNPFQFFGDDSPFCQDGS 88  
QY 98 PFYEYFKRLVPNPETIQEERADGG-----LNFGSGFII-SKNGYILTNTHVHVMG 148  
Db 89 PF-----QNSPFC--QGGNGNGGQQQKFMALGSGVVIDAAKGVVVTNNHVDNAS 138

QY 149 SIKVLNDKREYTAKLIGSDVQSVALLKIDATEELPWWKIGNPKNLKPGEWAAIGAPF 208  
Db 139 VIKVOLSDGRKFDAKVVGKDPDSIALIQIONPKNLTAIKLADSDALRVGDTTVAIGNPF 198  
QY 209 GFDSNVTAGIVSAKGRS-LPNESYTPPIQTDVAINPNSGGPLPNLKGQVVGINSQIYSR 267  
Db 199 GLGETVTSGIVSALGRSGLNVENFQTDAAINRSGALVNLNGELIGINTAILAP 258  
QY 268 SGFMGISFAIPIDVAMNVAEQLKNTCKVORGOLGVIIQEVSVYGLAQSFGLDKASGALIA 327  
Db 259 DGNIGIGFAIPSNMVKNLTSQMVYEQVGRKGLGIMGTELNSLAKAMKVDQAGAPVS 318  
QY 328 KILPGSPAERAGLQAGDIIVLSLDGGEIRSSGDLPMVMYGALTTPGKEVSLGVWRKGEETIK 387  
Db 319 QVMPNSSAAKAGIKAGDVITSLNGKPISSFAALRAQVGTMPVGSKISLGLLREGKAITVN 378  
QY 388 AKLGNAEHTGASSKTDAPYTEQSGTFSVES-----GITPLOTHTDSSGKHLVVVR 440  
Db 379 LEL-----QSSQSQVDSSTIFSGIEGAEKMSKNGQDKGVVWSVK 418  
QY 441 VSDAERAGLRHGDIELAVRASP 463  
Db 419 ANSPAAQIGLKKGDVIGANQQP 441

## RESULT 10

US-08-463-875A-2  
; Sequence 2, Application US/08463875A  
; Patent No. 5980907  
; GENERAL INFORMATION:  
; APPLICANT: DOUGAN, Gordon  
; APPLICANT: CHARLES, Ian G.  
; APPLICANT: HORMACHE, Carlos E.  
; APPLICANT: JOHNSON, Kevin S.  
; APPLICANT: CHATFIELD, Steven N.  
; TITLE OF INVENTION: LIVE VACCINES  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON and VANDERHVE PC  
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,875A  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/340,741  
; FILING DATE: 07-DEC-1994  
; APPLICATION NUMBER: US 07/952,737  
; FILING DATE: 30-NOV-1992  
; APPLICATION NUMBER: GB 9007194.5  
; FILING DATE: 30-MAR-1990  
; APPLICATION NUMBER: PCT/GB91/00484  
; FILING DATE: 28-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILSON, MARY J.  
; REGISTRATION NUMBER: 32,955  
; REFERENCE/DOCKET NUMBER: 117-158  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 475 amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-463-875A-2

```

Query Match 26.9%; Score 633.5; DB 2; Length 475;  
Best Local Similarity 34.1%; Pred. No. 7e-51;  
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;

```

RESULT 11
US-08-278-091-2
; Sequence 2, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24 973
: REFERENCE/DOCKET NUMBER: 1038-371
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 463 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-478-091-2

```

Query Match	26.9%;	Score 632;	DB 1;	Length 463;
Best Local Similarity	38.5%;	Pred. No. 9.3e-51;		
Matches 160;	Conservative 75;	Mismatches 153;	Indels 28;	Gaps 12;
QY	54	LPDFAQLVQSEGPAVVNIQIAPAPRTONGSGNAETDS-DPLADSDP--FYEFF-KRLVPN	109	
Db	28	LPFEVSEQNSLAEMLEKVQ--PAVTVLSVGKAKVDSPFLDDIPEEFKFFFGDRFAEQ	85	
QY	110	MPEIPOEADDDGFLNFGSGRII-SKNGYILTNTHVWAGMSIKVLLNDKREYAKTLIGSD	168	
Db	86	FGRGESKRNFRL--GSGVIIWASKGYVLTNNHIVTDEADKIYVQLQDGFREFKAKLVGD	143	
QY	169	VQSDVALLKDATAEELPVVKIGKNPKNLKPEWVAATGAPFGFONSVTAGIVSAKGRSLPN	228	
Db	144	EUSDIALVQLKPSNTEKTFADSDRLVGDFTVAIGNPFLGTQTVTSGIVSALGRSTGS	203	
QY	229	ES--YTPFTQTDVAINPGNSGGPLNFKGVGVGINSQIVSRSGFGWGISPAIPDIVAMNV	286	
Db	204	DSQTYENVITQDAAVNRKNSGGALVNLGELIGINTAIISPSCGNAGIAFAIPSNQASNL	263	
QY	287	ABQLKNTGKVRQGLGVITQEVSYGLAQSFGLDKASGALIIAKTLPGSPAEAGLQAGDIY	346	
Db	264	VQIILIEFGVRRLLGIKGELNADLAKAFNYSAQOGAFVSEVLPKSAAEKAGLKAGDII	323	
QY	347	LSLDGGETRSRSGDLPMVWGALTFGKEVSLGVWRKGBEITKAKLGNAAEHTGASSKDEA	406	
Db	324	TAMNGQKISSFAEIRAKTATTGAGKEISLTYLRDCKSHDVKKML-QADSSQLSSKT-EL	381	
QY	407	PYTEQSQSGTFSVESAGITLQTHDSSGRHLVVVR---SDAAERAGLRHGDIELAV	459	
Db	382	PALD-----GATLKDYDAKGVKGIEITTKTPNSLAAQR-GLKSGDIIIG	425	

```

RESULT 12
US-08-483-859-2
; Sequence 2, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/483,859  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/296,149  
;; FILING DATE: 26-AUG-1994  
;; APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/278,091  
;; FILING DATE: 21-JUL-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stewart, Michael I.  
;; REGISTRATION NUMBER: 24,973  
;; REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 595-1155  
;; TELEFAX: (416) 595-1163  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 463 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-483-859-2

Query Match 26.9%; Score 632; DB 1; Length 463;  
Best Local Similarity 38.5%; Pred. No. 9.3e-51;  
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

QY 54 LPDFAQLVQSEGPVAVNIQAAPRTQNGSNAETDS--DPLADSDP--FYEFF-KRLVNP 109  
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVTVLSVEGKAKVDSPFLDDIPEEFKFFGDRFAEQ 85  
QY 110 MPEIPEEADDGGLNFGSGFII--SKNGYILTNTHTVWAGMSIKVLLNDKREYTKALIGSD 168  
DB 86 FGGRGESKRNFRGL--GSGVIINASKGYVLTNNHVIDEADKITVQLQDGRFAKLVGKD 143  
QY 169 VQSDVALLKIDATELPPVYKIGNPKNLKPGEWAAIAGPFGDINSVTAGIVSAKGRSLPN 228  
DB 144 ELSDIALVQLEKPSNLTEIKFADSKLRVGDFTVAIGNPFGLGQTVTSGIVSALGRSTGS 203  
QY 229 ES--YTPFIOTDVAINPNSGGPLFNKGOVVGINSQIYSRSGFGMISFAIPIDVAMNV 286  
DB 204 DSGTYENYIOTDAAVNRNGSGGALVNLGELIGINTAIISPSGGNAGIAFAIPSNQASNL 263  
QY 287 AEQLKNTGKVRQGLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346  
DB 264 VQIILEFGVRRGLLGKIGKELNADLAKAFNVSAQQAFVSEVLPKSAAEKAGLKAGDII 323  
QY 347 LSLDGGEIRSSGDLPLVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAEHTGASSKTDEA 406  
DB 324 TAMNGOKISSFAEIRAKIATTGAGKEISLTLYLRDGKSHDVKMKL-QADDSSQLSSKT-EL 381  
QY 407 PYTEQSGTSTVESAGITLTQHTDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
DB 382 PALD-----GATLKDYDAKGVKGEITIKIOPNSLAAQR-GLKSGDIIIGI 425

RESULT 13  
US-08-472-173-2  
; Sequence 2, Application US/08472173  
; Patent No. 5665353  
; GENERAL INFORMATION:  
; APPLICANT: LOOMORE, Sheena M  
; APPLICANT: YANG, Yan-ping  
; APPLICANT: CHONG, Pele  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: KLEIN, Michael H.  
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
; TITLE OF INVENTION: Reduced Protease Activity  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Sim & McBurney  
;; STREET: Suite 701, 330 University Avenue  
;; CITY: Toronto  
;; STATE: Ontario  
;; COUNTRY: Canada  
;; ZIP: M5G 1R7  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/472,173  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/296,149  
;; FILING DATE: 26-AUG-1994  
;; APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/278,091  
;; FILING DATE: 21-JUL-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stewart, Michael I.  
;; REGISTRATION NUMBER: 24,973  
;; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 595-1155  
;; TELEFAX: (416) 595-1163  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 463 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-472-173-2

Query Match 26.9%; Score 632; DB 1; Length 463;

Best Local Similarity 38.5%; Pred. No. 9.3e-51;

Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

QY 54 LPDFAQLVQSEGPVAVNIQAAPRTQNGSNAETDS--DPLADSDP--FYEFF-KRLVNP 109  
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVTVLSVEGKAKVDSPFLDDIPEEFKFFGDRFAEQ 85  
QY 110 MPEIPEEADDGGLNFGSGFII--SKNGYILTNTHTVWAGMSIKVLLNDKREYTKALIGSD 168  
DB 86 FGGRGESKRNFRGL--GSGVIINASKGYVLTNNHVIDEADKITVQLQDGRFAKLVGKD 143  
QY 169 VQSDVALLKIDATELPPVYKIGNPKNLKPGEWAAIAGPFGDINSVTAGIVSAKGRSLPN 228  
DB 144 ELSDIALVQLEKPSNLTEIKFADSKLRVGDFTVAIGNPFGLGQTVTSGIVSALGRSTGS 203  
QY 229 ES--YTPFIOTDVAINPNSGGPLFNKGOVVGINSQIYSRSGFGMISFAIPIDVAMNV 286  
DB 204 DSGTYENYIOTDAAVNRNGSGGALVNLGELIGINTAIISPSGGNAGIAFAIPSNQASNL 263  
QY 287 AEQLKNTGKVRQGLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346  
DB 264 VQIILEFGVRRGLLGKIGKELNADLAKAFNVSAQQAFVSEVLPKSAAEKAGLKAGDII 323  
QY 347 LSLDGGEIRSSGDLPLVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAEHTGASSKTDEA 406  
DB 324 TAMNGOKISSFAEIRAKIATTGAGKEISLTLYLRDGKSHDVKMKL-QADDSSQLSSKT-EL 381  
QY 407 PYTEQSGTSTVESAGITLTQHTDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
DB 382 PALD-----GATLKDYDAKGVKGEITIKIOPNSLAAQR-GLKSGDIIIGI 425

RESULT 14  
US-08-487-167-2  
; Sequence 2, Application US/08487167

Patent No. 5869302  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M.  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
REDUCED PROTEASE ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487.167  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296.149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278.091  
FILING DATE: 21-JUL-1994  
NAME: Stewart, Michael I.  
ATTORNEY/AGENT INFORMATION:  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-167-2

Query Match 26.9%; Score 632; DB 2; Length 463;  
Best Local Similarity 38.5%; Pred. No. 9.3e-51;  
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

Qy 54 LPDFAQLVQSEGPVAVNIQAAPRTQNGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109  
Db 28 LPSFVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSRSPFLDDIPPEPKFFGDFAEQ 85

Qy 110 MPEIQEADGGNFGSGFII-SKNGVILTNTHVAGMGSIKVLLNDKREYTKLIGSD 168  
Db 86 FGGRGESKRNRGL--GSGVIINASKGYVLTNNHVIDEADKITVQLQDGRFKAKLVGKD 143

Qy 169 VQSDVALLKIDATELPPVVKIGNPNKLPGEWVAIGAIPFGFDSNVTAGIVSAKGRSLPN 228  
Db 144 ELSDIALVQLEKPSNLTEIKFADSKLRVGDFTVAIGNPFGLGQTVTSIGIVSALGRSTGS 203

Qy 229 ES--YTPFIQTDVAIPNCGSGPLNKGQVVGINSOIVSRSGGFMGISFAIPIDVAMNV 286  
Db 204 DSGYENYIQTDAVNRNCGSALVNLGELIGINTAIIPSGCGNAGIAFAPISQNASNL 263

Qy 287 AEQLKNTQKVRQGLVITQIYESYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346  
Db 264 VQOILEFGVRRGLIGKGGELNADLAKAFNVSAQOAGFVSEVLPKSAEAKGLKAGDII 323

Qy 347 LSLDGGEIRSSGDLPMVGAITPGKEVSLGVVRKGEETIKAKLGNAAEHTGASSKTDEA 406

Db 324 TAMNGQISSFAEIRAKIATTGAGKEISLTYLROGKSHDVKNKL-QADDSSQLSKT-EL 381  
Qy 407 PYTEQOQSTSVESAGITLQHTDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
Db 382 PALD-----GATLKDYDAKGVKGIEITRIKTOPNSLAAQR-GLKSGDIIIGI 425

RESULT 15  
US-08-482-816-2  
Sequence 2: Application US/08482816  
Patent No. 5935573  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease  
ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482.816  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296.149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278.091  
FILING DATE: 21-JUL-1994  
NAME: Stewart, Michael I  
ATTORNEY/AGENT INFORMATION:  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-482-816-2

Query Match 26.9%; Score 632; DB 2; Length 463;  
Best Local Similarity 38.5%; Pred. No. 9.3e-51;  
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

Qy 54 LPDFAQLVQSEGPVAVNIQAAPRTQNGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109  
Db 28 LPSFVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSRSPFLDDIPPEPKFFGDFAEQ 85

Qy 110 MPEIQEADGGNFGSGFII-SKNGVILTNTHVAGMGSIKVLLNDKREYTKLIGSD 168  
Db 86 FGGRGESKRNRGL--GSGVIINASKGYVLTNNHVIDEADKITVQLQDGRFKAKLVGKD 143

Qy 169 VQSDVALLKIDATELPPVVKIGNPNKLPGEWVAIGAIPFGFDSNVTAGIVSAKGRSLPN 228  
Db 144 ELSDIALVQLEKPSNLTEIKFADSKLRVGDFTVAIGNPFGLGQTVTSIGIVSALGRSTGS 203



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:21 ; Search time 20.95 Seconds  
(without alignments)  
2132.770 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKQYFALAALCAALLAG.....ERAGLRHGDEILAVRASPRQ 465  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2246	95.5	499	2 B81914	probable periplasm
2	985.5	41.9	474	2 F83550	serine proteinase
3	862.5	36.7	514	2 A82581	periplasmic protei
4	735	31.2	523	2 A97479	probable serine pr
5	735	31.2	523	2 A12696	serine proteinase
6	716	30.4	493	2 C97605	probable serine pr
7	716	30.4	514	2 A82827	serine proteinase
8	693.5	29.5	513	2 AD3418	serine proteinase
9	692.5	29.4	513	2 I40060	serine proteinase
10	683	29.0	497	2 B81728	serine proteinase,
11	683	29.0	497	2 H71465	probable do serine
12	683	29.0	524	2 AG3328	proteinase do (EC
13	669.5	28.5	488	2 G81528	serine proteinase,
14	668.5	28.4	488	2 H86612	DO serine proteina
15	668.5	28.4	488	2 G72011	do serine proteina
16	653.5	27.8	457	2 AG0433	proteinase (EC 3.4
17	649	27.6	481	2 AH0410	global stress requ
18	648	27.5	466	1 A64113	heat shock protein
19	639	27.2	474	2 A13349	proteinase DO (EC
20	638	27.1	511	2 F97720	periplasmic serine
21	637	27.1	474	1 I40059	htrA-like protein
22	636.5	27.1	475	2 AC0528	protease DO precu
23	636	27.0	530	2 F87590	serine proteinase
24	633.5	26.9	475	1 S15337	heat shock protein
25	632.5	26.9	456	2 F82307	proteinase DO VC05
26	631	26.8	429	2 AD1894	serine proteinase
27	627.5	26.7	455	2 JC6051	trypsin-like prote
28	623.5	26.5	455	2 C91142	serine endoprotein
29	617.5	26.2	455	2 F85987	serine endoprotein

30	617.5	26.2	455	2 AB0909	serine protease (E
31	615.5	26.2	513	2 B71722	probable periplasm
32	614.5	26.1	459	2 F72359	periplasmic serine
33	610.5	25.9	474	2 S45229	proteinase DO (EC
34	610.5	25.9	474	2 E85500	proteinase DO (EC
35	610.5	25.9	474	2 E90849	serine proteinase
36	606.5	25.8	452	2 S77538	heat shock protein
37	604.5	25.7	481	2 D82826	proteinase hhoB (E
38	600.5	25.5	416	2 S75445	proteinase do prec
39	594.5	25.3	478	2 G84956	proteinase do (EC
40	589	25.0	389	2 B83089	AlpW protein PA444
41	582	24.7	476	2 H71936	proteinase DO - He
42	574.5	24.4	443	1 C64647	serine proteinase
43	571	24.3	453	1 B70426	periplasmic serine
44	564	24.0	468	2 AI2811	serine proteinase
45	564	24.0	495	2 B97590	htrA protein homol

ALIGNMENTS

RESULT 1

B81914

probable periplasmic serine proteinase (EC 3.4.21.-) NMA0710 [imported] - Neisseria

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: B81914

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.;

: Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 224

A:Reference number: A81775; MUID:20222556

A:Accession: B81914

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-499 <PAR>

A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB83996.1; PID:g7

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: NMA0710

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

Query Match 95.5%; Score 2246; DB 2; Length 499;

Best Local Similarity 95.9%; Pred. No. 4.6e-138;

Matches 444; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 VFKKQYFALAALCAALLAGCEKAGSFFGADKKKASVFVERIEHTKDDGSVSMLLPDFVQL 60

DB 1 MFKKQYLALALCAASLAGCDKAGSFFGADKKKASVFVERIKHTKDDGSVSMLLPDFVQL 60

QY 61 VQSEGPAVNIQAPAPRTQNGSNAETSDPLADSDPFYEFFKRLVPMNPEIPOEADD 120

DB 61 VQSEGPAVNIQAPAPRTQNGSNAETSDPLADSDPFYEFFKRLVPMNPEIPOEADD 120

QY 121 GGLNFGSGFIISKNGYILTNTHVVVAGMGSIKVLLNDKREYTKAKLIGSDVQSDVALLKIDA 180

DB 121 GGLNFGSGFIISKNGYILTNTHVVVAGMGSIKVLLNDKREYTKAKLIGSDVQSDVALLKIDA 180

QY 181 TEELPVVYKIGNPKNKLKGEVWAAIGAPGFDNSVTAGIVSAKGRSLPNESYTPFTQDVA 240

DB 181 TEELPVVYKIGNPKDLKPGVWAAIGAPGFDNSVTAGIVSAKGRSLPNESYTPFTQDVA 240

QY 241 INPNSGGPLFNKLGQVVGINSQIYSRSGGFMGISFATPIDVAMVAFOLKNTGVQRGQ 300

DB 241 INPNSGGPLFNKLGQVVGINSQIYSRSGGFMGISFATPIDVAMVAFOLKNTGVQRGQ 300

QY 301 LGVIIQEVSVYGLAQSGFLDKASGALIAKILPGSPAERAGLAGDVLVSLDGGEIRSSGDL 360

DB 301 LGVIIQEVSVYGLAQSGFLDKAGGALIAKILPGSPAERAGLAGDVLVSLDGGEIRSSGDL 360

QY 361 PWMVGAITPGKEVSLGVWRKGEITIRAKLGNAAEHTGASSKTDEAPYTEQOOSGTFSVES 420

||||| 361 PVMVGATPQKEYSLGVRKGEETIKVLCGNAAEHIGASSKTDAPYTEQQSGTFVES 420  
QY 421 AGITLTQHTDSSGKHLVVVSVDAERAGLRHGDDEILAVRASP 463  
Db 421 AGITLTQHTDSSGHLVVVSVDAERAGLRHGDDEILAVGQVP 463

RESULT 2  
F83550  
serine proteinase MucD precursor PA0766 [imported] - Pseudomonas aeruginosa (strain PA01  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83550  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim  
; Loy, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
gen  
Reference number: A89950; MUID:20437337  
Accession: F83550  
C:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <STO>  
A:Cross-references: GB:AE004511; GB:AE004091; NID:99946646; PIDN:AAG04155.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: mucD; PA0766

Query Match 41.9%; Score 985.5; DB 2; Length 474;  
Best Local Similarity 52.5%; Pred. No. 2.1e-56;  
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

QY 54 LPDFAOLVOSEGPVAVNI---QAAPRTQNGSGNAETDSDPLADSDP-FYEFKRLVFN 109  
Db 27 LPDFTPLVEQASPAVNISTRQKLP----DRAMARGOLSTPDLEGLPMPFRDFLERSIPQ 82

QY 110 MPEIP---QEEADGGLNGSGFIISKNGYILTNTHVVGAGMSIKVLLNDKREYTAKLIG 166  
Db 83 VPRNPRGQREQA---SLGSGFIISNDGYILTNHVVADADETLVRLSDRSEHAKKLIG 138

QY 167 SDVQSDVALLKIDATELPPVVKIGNPKNKPGEWVAAGPFGFDSNVTAGIVSAKRSL 226  
Db 139 ADPRSDVAVLKIDR-KNLPTLKLGDSNKLKVGWVLAIGSPFGFDHVTAGIVSAKRSL 197

QY 227 PNESTYPTFQTDVAIPNGSGGFLNKGQVVGINSQIYSRSGFGMGSIFAIPIDVAMNV 286  
198 PNESTYPTFQTDVAIPNGSGGFLNKGQVVGINSQIYSRSGFGMGSIFAIPIDVAMNV 257

QY 287 ABQLKNTGKVRQGLGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346  
Db 258 ADQLKAGKVRSGVLGVITQEVNKNDAESFGLDKPSGALVAQLVEDGPAKGGVLQVGDVI 317

QY 347 LSLDGGEIRSSGDLPMVVGAITPGKEVSLGWRKGEETIKAKLGNAAEHTGASSKTD- 405  
Db 318 LSLUNGOSINESADPLHLVGNMKPKDINLDVIRNGQRKSLSMVAGSLPD-----DDEE 370

QY 406 -----APYTEQQSGTFSESAGITLQTH---TDSGSKHLVVVSVDA-AERAGLRHGDDEI 456  
Db 371 IASMGAPGAERSSNRLGVTVVADITAEQRKSLDIQG-GVVIKEVQDGPAAVIGLRPGDVI 428

RESULT 3  
A82581  
periplasmic proteinase XF2241 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: The Xylella  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: A82581  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-514 <SIM>  
A:Cross-references: GB:AE004037; GB:AE003849; NID:g9107394; PIDN:AAF85040.1; GSPDB:B  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Carre  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrato, D.M.; Carre  
as-Neto, E.; Decena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; F  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; L  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Marti  
A:Authors: Martins, A.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2241  
C:Superfamily: Helicobacter serine proteinase

Query Match 36.7%; Score 862.5; DB 2; Length 514;  
Best Local Similarity 46.2%; Pred. No. 2.2e-48;  
Matches 194; Conservative 64; Mismatches 129; Indels 33; Gaps 10;

QY 54 LPDFAOLVOSEGPVAVNIQAAPRTQNGSGNAETDSDPLADSDPYPYEFKRL-----V 107  
Db 54 LPDFTQLVDQVGPVVNIETVITRKVKGRRIPLND-----IPEFPRRFFGPDFQM 106

QY 108 PNWPEIPEEADGGGL---NFGSGFIISKNGYILTNTHVVGAGMSIKVLLNDKREYTAKL 164  
Db 107 PNQPRGGQD---DEGGTAGRGMSGGFIISKDGYILTNHVTIGTASEVTIKUTDRREKAKI 164

QY 165 IGSQSDVALLKIDATELPPVVKIGNPKNKPGEWVAAGPFGFDSNVTAGIVSAKR 224  
Db 165 IGSDEQYDVALKIDR-KNLPTVRIGDSSLSKSGQWVAIGSPFGLDHSVTAGIVSALGR 223

QY 225 SLPNES-YTPFTQTDVAIPNGSGGFLNKGQVVGINSQIYSRSGFGMGSIFAIPIDVA 283  
Db 224 STSDDRYVPFTQTDVPIQNGSGGFLNTRGEVIGINSQIFSASGSGMGSIFAIPINLA 283

QY 284 MNVAEQKNTGKVRQGLGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAG 343  
Db 284 INAAEQIRKTKVQRSMGLGVETGIDALKAQGLGLPDSRCALVNNIPPHSPAAGAGIEVC 343

QY 344 DIVLSLDGGEIRSSGDLPMVVGAITPGKEVSLGWRKGE--ETIKAKLGNAAEHTGASS 401  
Db 344 DVIRSVNGKVISFSFDLPPLICMPPGPKATLGIIRDGKPREIVTWLTSALN-QDATASED 402

QY 402 KTDAPYTEQQSGTFSESAGITLQTHDSSSKHLVVVSVDAERAGLRHGDDEILAVRA 461  
Db 403 EDDTAPSKPETSQSA--NVELLGLQVNLSSAERERL-----ASSQNAKGGVRIITAVTA 452

RESULT 4  
A97479  
probable serine proteinase homolog precursor [imported] - Agrobacterium tumefaciens  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: A97479  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourollo, B.; Gol  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
A:Reference number: A97359; PMID:11743194  
A:Accession: A97479  
A:Status: preliminary  
A:Molecule type: DNA







C:Accession: I40060  
R:Tatum, F.M.; Cheville, N.F.; Morfitt, D.  
Microb. Pathog. 17, 23-36, 1994  
A:Title: Cloning, characterization and construction of htrA and htrA-like mutants of Br  
A:Reference number: I40059; MUID:95165990  
A:Accession: I40060  
A:Molecule type: DNA  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-513 <RES>  
A:Cross-references: EMBL:U07352; NID:g497156; PIDN:AAA70164.1; PID:g497157  
C:Genetics:  
A:Gene: htrA  
C:Function:  
A:Description: EC 3.4.21.-; serine proteinase [validated, MUID:95165990]; permits growth  
C:Superfamily: Helicobacter serine proteinase  
C:Keywords: hydrolase; serine proteinase

Query Match 29.4%; Score 692.5; DB 2; Length 513;  
Best Local Similarity 37.4%; Pred. No. 2.3e-37;  
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

QY 9 ALAALCAALLGCEKAGSEFFGADKKKAESEFVERIEHTKDDGSVSM--LPPD-----FAQLVQSEG 65  
DB 14 AAVALSAL-----AGAFVVTGPGALNEARAE-----AVHTVPPAGFADLVKVR 61  
QY 66 PAVNIIQAAPARTQNGSGN---AETSDPLADSDPFYEFF-----KRLV 107  
DB 62 PAVSVRVKDVQETSNRGPPOFFGPPGFDLPDGHPLKRFRRDFGMEPRGDSRSDNRGK 121  
QY 108 PNMPEIPEEADDGLNFGSGFIISKNGYILTNTHVAGMGSIKVLNDRKREYAKLIGS 167  
DB 122 ANKPRPGHERP-----VAOGSGFVISEDGVVYVNNHWSGDGAYTVVLDGTDELAKLIG 177  
QY 168 DVQSDVALLKIDATE-ELPVVKVIGKPNLKPGEVAAIGAPFGFNSVTAGIVSAKGRSL 226  
DB 178 DPTDVLVLLKNAPKREKVVYAFGDNDKVRVGDVVVAVNGNPFGLGTVTSGIVSAGRDI 237  
QY 227 PNESTPTPTQTDVAINPNSGGPLNLKQGVVINSQIYRSRGGFMGIFSFAIPIDVAMNY 286  
DB 238 GAGPYDDPFIQIDAAVNKNSGGPAPDLSEVIGINTAIFSPSGSVGIAFAIPSSAKQV 297  
QY 287 AEOLKNTKVGORGOLGVLIOEVSYSYLAQSLDGLKAGALIAKILPGSPAERAGLQGDIV 346  
DB 298 VDQLIKGSEVBERGWIGVQIQPTKDIAASLGAEKGAIVASPDGPAKAGIKAGDVI 357  
QY 347 LSLDGEIRSSGDLPMVVGATTPGKEVSLGVWR--KGEITIK-AKLGNAAEHTGASSKT 403  
DB 358 TAVNGETVODPRDLARKVANTAPGEKAALTVMRKNKAEINVTIAAMPNDKKGSGSQSND 417  
QY 404 DEAPYTEQSGTFSVESAGITLQTHDSSGKHLVVVRV---SDAAERAGLRHGDILAV 459  
DB 418 NDGGQGE-----TLDSYGLTVWPSD--GKGWVTVDPDPSDAADR-GIRSGDVIVSV 467

RESULT 10  
B81728  
serine proteinase, HtrA/DegQ/DegS family TC0210 [imported] - Chlamydia muridarum (strain  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 21-Jul-2000  
C:Accession: B81728  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: AB1500; MUID:20150255  
A:Accession: B81728  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-497 <TET>  
A:Cross-references: GB:AB002288; GB:AE002160; NID:g7190247; PIDN:AAF39082.1; PID:g719024  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:

A:Gene: TC0210  
C:Superfamily: Helicobacter serine proteinase

Query Match 29.0%; Score 683; DB 2; Length 497;  
Best Local Similarity 37.4%; Pred. No. 9e-37;  
Matches 173; Conservative 77; Mismatches 154; Indels 58; Gaps 11;

QY 28 FGADKKKAEASFVERIEHTKDDGSVSM--LLPD-----FAQLVQSEGPAVVNIOAAPRTQ- 80  
DB 22 YSAPKKDSSTGICLAASQSDRELQEDLLKEVSRGFSKVAQAATPGVVIENFPKTSOA 81  
QY 81 -----NGSGNAETSDPLADSDPFYEFFKRL--VPNMPEIPEEADDGLNFGSGFIISK 133  
DB 82 IASPGNKRGFQENPFDFND-----EFNRFGLPSHREQRPQORDAVR--GTGFIIVSE 134  
QY 134 NGYILTNTNVAGMSIKVLNDRKREYAKLIGSDVQSDVALLKIDATEELPVVIGNPK 193  
DB 135 DGIVVTHVHVEDAGKIHVTLHDGOKYAKIIGLDPKTDLAVIKIOA-KNLFLFTFGNSD 193  
QY 194 NLKPGEWVAAGAPFGFNSVTAGIVSAKGRS-LPNESYTPFIOTDVAINPNSGGPLFN 252  
DB 194 QLOIGDMSIAIGNPFGLOATVTVGVISAKGRNLHIVDFEDFIQTDAAINPNSGGPLN 253  
QY 253 LKGVGVINSQIYRSRGGFMGIFSFAIPIDVAMNVAELKNTGKVGORGOLGVIIEVSYGL 312  
DB 254 IDGVVIGVNTAIVSGSGYIGIGFAIPSLMAKRVTDQLTSDGOVTRGVLGTLPIDSEL 313  
QY 313 AQSFGLDKASCALLAKILPGSPAERAGLQGDIVLSLDGGEIRSSGDLPMVVGATTPGKE 372  
DB 314 AACYKLEKVVYVYVAGLITDVVYKSGSPAERAGLQEDIVVAYNGKEVESLALRNAISLMMPPGR 373  
QY 373 VSLGVWRKGEITIKAKLGNAAEHTGASSKTDEAPYTSQ-----SGTFSVESAGITLQPH 428  
DB 374 VLKVVREGRFI-----EIPVTVTQIPAEQGVSAQKMGVRVQNL 413  
QY 429 T-----DSSGKHLVVVRVSDAERAGLRHGDILAV 459  
DB 414 TPECKKLGLASDTRGIFVVSVEAGSPAASAGVVPQILAV 455

RESULT 11  
H71465  
probable do serine proteinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitc  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia  
A:Reference number: A71570; MUID:99000809  
A:Accession: H71465  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-497 <ARN>  
A:Cross-references: GB:AE001355; GB:AE001273; NID:g3329292; PIDN:AAC68420.1; PID:g3  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: htrA  
C:Superfamily: Helicobacter serine proteinase

Query Match 29.0%; Score 683; DB 2; Length 497;  
Best Local Similarity 37.0%; Pred. No. 9e-37;  
Matches 171; Conservative 75; Mismatches 158; Indels 58; Gaps 11;

QY 28 FGADKKKAEASFVERIEHTKDDGSVSM--LLPD-----FAQLVQSEGPAVVNIOAAPRTQ- 80  
DB 22 YSAKKDSKADICLAVSSGQEDLLKEVSRGFSRVAAKATPGVVIENFPKTSOA 81  
QY 81 -----NGSGNAETSDPLADSDPFYEFFKRL--VPNMPEIPEEADDGLNFGSGFIISK 133  
DB 82 IASPGNKRGFQENPFDFND-----EFNRFGLPSHRE--QORPQORDAVRGTGFIIVSE 134



[illegible]

RESULT 15  
G72011 do serine proteinase - Chlamydoiphila pneumoniae (strain CWL029)  
C:Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: G72011  
R:Kalmán, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: G72011  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-488 <ARN>  
A:Cross-references: GB:AE001678; GB:AE001363; NID:g4377301; PIDN:AAD19116.1  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: htrA  
C:Superfamily: Helicobacter serine proteinase

Search completed: June 10, 2002, 12:09:52  
Job time: 31 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:10:31 ; Search time 13.48 seconds  
(without alignments)  
1335.652 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQYFALAALCAALLAG.....ERAGLRHGDEILAVRASPRQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	747	31.7	504	1	DEGP_RHIME
2	731.5	31.1	503	1	DEGP_BARHE
3	692.5	29.4	513	1	DEGP_BRUB
4	683	29.0	497	1	DEGP_CHLMU
5	683	29.0	497	1	DEGP_CHLTR
6	668.5	28.4	488	1	DEGP_CHLPN
7	648	27.5	466	1	HTOA_HAEIN
8	633.5	26.9	475	1	DEGP_SALTY
9	627.5	26.7	455	1	DEGP_RICPR
10	615.5	26.2	513	1	DEGP_ECOLI
11	610.5	25.9	474	1	DEGP_ECOLI
12	594.5	25.3	478	1	DEGP_BUCAI
13	571.5	24.3	478	1	DEGP_BUCAP
14	470	20.0	437	1	DEGP_ARATH
15	448	19.0	480	1	HRA1_HUMAN
16	447.5	19.0	355	1	DEGS_ECOLI
17	436.5	18.6	413	1	HTRA_LACHE
18	435.5	18.5	480	1	HRA1_MOUSE
19	435	18.5	448	1	DEGS_ARATH
20	428	18.2	340	1	DEGS_HAEIN
21	414.5	17.6	458	1	HRA2_HUMAN
22	412.5	17.5	476	1	HRA4_HUMAN
23	412	17.5	400	1	YXAX_BACSU
24	410	17.4	458	1	HRA2_MOUSE
25	388	16.5	408	1	HTRA_LACLA
26	353	15.0	460	1	HRA3_MOUSE
27	344.5	14.6	321	1	HROA_ARATH
28	303.5	12.9	452	1	HRA3_HUMAN
29	230	9.8	630	1	Y4BJ_RHISN
30	198.5	8.4	997	1	YNN3_YEAST
31	137.5	5.8	767	1	DLG4_HUMAN
32	131	5.6	724	1	DLG4_RAT
33	130	5.5	724	1	DLG4_MOUSE

34	128	5.4	2021	1	OMPA_RICCN	Q52657 rickettsia
35	118	5.0	280	1	ETA_STAAU	P09331 staphylococ
36	115.5	4.9	837	1	APBL_HUMAN	Q02410 homo sapien
37	115	4.9	568	1	YELL_DROMD	Q09p71 drosophila
38	115	4.9	568	1	YELL_DROSU	Q02437 drosophila
39	114.5	4.9	931	1	DLGI_RAT	Q62696 rattus norv
40	114	4.8	904	1	DLGI_HUMAN	Q12959 homo sapien
41	114	4.8	2249	1	OMPA_RICRI	P15921 rickettsia
42	113.5	4.8	2364	1	PGCA_BOVIN	P13608 bos taurus
43	113	4.8	545	1	CH60_CAMJE	O69289 campylobact
44	111	4.7	261	1	DER3_DERPT	P39675 dermatophag
45	110.5	4.7	588	1	RUBB_BRANA	P21241 brassica na

## ALIGNMENTS

RESULT 1  
DEGP\_RHIME  
ID DEGP\_RHIME STANDARD; PRT; 504 AA.  
AC Q52894;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Probable serine protease do-like precursor (EC 3.4.21.-).  
GN DEGP1 OR DEGP OR R01021 OR SMC02365.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=961146524; PubMed=8550509;  
RA Glazebrook J., Ichige A., Walker G.C.;  
RT "Genetic analysis of Rhizobium meliloti bacA-phoA fusion results in  
RT identification of degP: two loci required for symbiosis are closely  
RT linked to degP.";  
RL J. Bacteriol. 178:745-752(1996).  
[2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
RA "Analysis of the chromosome sequence of the legume symbiont  
RA Sinorhizobium meliloti strain 1021.";  
Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
CC DEGP/DEGP/DEGS FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U31512; AAC43669.1; ALT\_INIT.  
CC EMBL; AL591785; CAC45593.1; -.  
CC MEROPS: S01.273; -.  
CC DR InterPro; IPR001478; PDZ.  
CC DR InterPro; IPR001940; Protease2C.  
CC DR InterPro; IPR000126; Ser\_proteas\_V8.  
CC DR InterPro; IPR001254; Trypsin.  
CC DR Pfam; PF00595; PDZ; 2.  
CC DR Pfam; PF00089; trypsin; 1.

[illegible]



Db 302 LGVQTPVTKELSDSIGLKEAKGALITDPLKG-PAKAGIKAGDVIIISVNGEIKINDVROL 360  
Qy 361 PVMGCAITPGREVLGVRKGEITIKAKLGNAAEHTGASSKTDEAPYT-EQSGTFSVE 419  
Db 361 AKRIANMSPGETVTLGVWKSKEENIKVKLDSMPED---ENMKDGSKYSNEHCNSDETLE 417  
Qy 420 SAGITLQTHDTSSGKHLVVVVR---SDAERAGLRHGRBILAV 459  
Db 418 DYGLIVAPSDGCG--LVVTDVDPDSDAADK-GIRPGDVIVTV 457

RESULT 3  
DEGP\_BRUAB  
ID DEGP\_BRUAB STANDARD; PRT: 513 AA.  
AC Q44597;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable serine protease do-like precursor (EC 3.4.21.-).  
GN DEGP OR HTRA.  
Bacteria abortus.  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Brucellaceae; Brucella.  
NCBI\_TaxID=235;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=2308;  
RX MEDLINE=95165990; PubMed=7861951;  
RA Tatsum F.M., Chevillon N.F., Morfitt D.;  
RT "Cloning, characterization and construction of htra and htra-like  
mutants of Brucella abortus and their survival in BALB/c mice.";  
RL Microb. Pathog. 17:23-36(1994).  
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
DEGP/DEGO/DEGS FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U07352; AAA70164.1; -  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR Pfam; PF00595; PDZ; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00834; PROTEASES2C.  
DR SMART; SM00228; PDZ; 2.  
DR PROSITE; PS0106; PDZ; 2.  
KW Hydrolase; Serine protease; Signal; Periplasmic.  
FT SIGNAL 1 25 POTENTIAL.  
FT CRAIN 26 513 PROBABLE SERINE PROTEASE DO-LIKE.  
FT DOMAIN 125 299 CATALYTIC.  
FT PDZ 1.  
FT DOMAIN 300 391 PDZ 2.  
FT DOMAIN 414 500 PDZ 1.  
FT ACT\_SITE 152 152 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 182 182 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 257 257 CHARGE RELAY SYSTEM (POTENTIAL).  
SQ SEQUENCE 513 AA; 53483 MW; DE1CEFF1959472806 CRC64;

Query Match 29.4%; Score 692.5; DB 1; Length 513;  
Best Local Similarity 37.48; Pred. No. 4e-37;  
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

Qy 9 AALACALLACGKAGSFFGADKKESFVERIETHKDDGSVMLLP---DFAQLVOSBG 65  
Db 14 AAVALSRAAL-----AGAFVVTGPIGALNEARAE-----AVHVTTPPQAGFADLVEKVR 61

Qy 66 PAVNIIQAAPAPRTONGSGN---AETDSPLADSDPPFYEFF-----KRIV 107  
Db 62 PAVSVVRVKDVQETSNRGQFPGPPGPDQLPDGHPKLRFRDFCMGPRGDSRSDNRCK 121  
Qy 108 PNMEIPEQERADDGGLNFGSGFIISKNGIYILTNTHVHVMGMSIKVLLNDKREYTKLIGS 167  
Db 122 ANKPRPGHERP---VAQSGFVISEDGVYVTVNNHVSVDGDAYTVLDDGTDLAKLIGA 177  
Qy 168 DVOSDVALLKIDATE-ELPVVKIGNPKNLKPGEWAAIGAPFGFNSVTAGTIVSAKGRSL 226  
Db 178 DPTDLDLAVLKINAPKRFVYVAFGDNDKVRVGDWVAVGNPFLGGTGTSGIVSARGDI 237  
Qy 227 PNESTPTFIQTDVAINFGNSGGPLFNKGVVGVINSQIYSRSGGFMGIFSFAIPIDVAMV 286  
Db 238 GAGPYDDFIQIDAANAANGNSGGPAPDLSEVIGINTAIFSPGSGSVGIAFAIPSTAKQV 297  
Qy 287 AEQLKNTGKVORGQLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346  
Db 298 VDQLIKKGSVERGWIGVQIPVTKDIAASGLAEKGAIVASPDQDGPAAKAGIKAGDVI 357  
Qy 347 LSLDGGIRSGGDLPMVMVGAIITPGKEVSLGVWR--KGEEITIK-AKLGNAAEHTGASSKT 403  
Db 358 TAVNGETVQDPRDLARKVANITAPGEKAALTVMRKNKABEINVTIAMPNDKSGSQSND 417  
Qy 404 DEAPYTEQSGTFSVESAGITLQTHDSSGKHLVVVVR---SDAERAGLRHGRBILAV 459  
Db 418 NDGSGE-----TLDVGLTVVPSD--GKGVVTVDPDSDAADR-GIRSGDVIVSV 467

RESULT 4  
DEGP\_CHLMU  
ID DEGP\_CHLMU STANDARD; PRT: 497 AA.  
AC Q9PL97;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable serine protease do-like precursor (EC 3.4.21.-).  
GN DEGP OR HTRA OR TC0210.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MoPn / Ni99;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
DEGP/DEGO/DEGS FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; A500228; AAF39082.1; -  
DR TIGR; TC0210; -  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR InterPro; PR000126; Ser.proteas\_v8.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00595; PDZ; 2.



Db 254 INGOVICNTAIVSGSGYIGIGFAIPSLMAKRVIDQLISDQVTRGFLGTQLQIDSEL 313  
QY 313 AQSGFLDKASGLAKTILPGSPAERAGLOAGDIVLSLDGGIRSSGDLPLVMVGAITPGKE 372  
Db 314 ATCYKLEKVGALVTDVVKSGPAKAGLRQEDIVVAYNGKEVESLSALRNALSLMMPGTR 373  
QY 373 VSLGVWRKGEITIKAKLGNAAHTGASSKTDEAPYTEQ-----SGTFSVESAGITLQTH 428  
Db 374 VLKIVREG-----KRIEIPVTVTQIPTEDGVSAQKMGVRVQNI 413  
QY 429 T-----DSSGKHLVVVRVSDAAERAGLRHGDEILAV 459  
Db 414 TPEICKKLGLAADTRGILWAVEAGSPAASGAVAPGQILAV 455  
RESULT 6  
ID DEGP\_CHLPN STANDARD; PRT; 488 AA.  
AC Q9Z6T0; Q9J0D7; Q9KIW4;  
DE 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE Probable serine protease do-like precursor (EC 3.4.21.-).  
GN DEGP OR HTRA OR CPN0979 OR CP0877;  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
R. Lingner L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
DEGP/DEGP/DEGS FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE001678; AAD19116.1; .  
DR EMBL; AE002246; AAF38665.1; .  
DR EMBL; AP002548; BAA99186.1; .  
DR MEROPS; S01.273; .

DR PHCI-2DPAGE; Q9Z6T0; .  
DR TIGR; CP0877; .  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00595; PDZ; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00834; PROTEASES2C.  
DR SMART; SM00228; PDZ; 2.  
DR PROSITE; PS0106; PDZ; 2.  
KW Hydrolase; Serine protease; Signal; Complete proteome.  
FT SIGNAL 1 20  
FT CHAIN 21 488  
FT DOMAIN 119 280  
FT DOMAIN 281 372  
FT DOMAIN 388 476  
FT ACT\_SITE 134 134  
FT ACT\_SITE 164 164  
FT ACT\_SITE 238 238  
SQ SEQUENCE 488 AA; 52311 MW; 0EE7E0F8F106F49 CRC64;  
Query Match 28.4%; Score 668.5; DB 1; Length 488;  
Best Local Similarity 36.1%; Pred. No. 1.3e-35;  
Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;  
QY 10 LAALCAALLAGCEKAGFFGADKKKEAFVERIETHKDDGSVSMLLPDPFAQLVQSEGPVAV 69  
Db 10 LAVLVGSSLLALPLSGQAVG--KKES---RVSELPQDVLKKEISGGFSKVATKATPAV 63  
QY 70 NQAAAPR--TONGSNAETSDPLADSDPFY-EFFKRL--VPNMPEIPO-EAADDGGL 123  
Db 64 YIESFKSQAVTHPSFCRRGYPENPF---DYFNDEFFNRFGLPSQREKPKQSKAVR--- 117  
QY 124 NFGSGFIISKNGYILNTHVWAGSGIKVLLNDKREYTKAKLGSDVSDVALLKTDAT 183  
Db 118 --GTGELVSPDGYIVTNHVVEDTKIHLVLDGCKPAPVIGLDPKTDLAVIKI-KSQN 174  
QY 184 LPVVKIGNPKNLKPGEMVAAGAPGFDNSVTAGIVSAKGRS-LPNESYTPFIOTDVAIN 242  
Db 175 LPVLFGNSDHLKVGWDWAIAGNPFGLQATVTVGVISAKGRNQLRIADFDFIQTDAAIN 234  
QY 243 PGNSGGLFLNLKQGVGINSQIYRSRSGFGMGISFAIPIDVAMNVAEOLKNTGKVQRGOLG 302  
Db 235 PGNSGGLPLNIDGQVIGVNTAIVSGSGYIGIGFAIPSLMANRIIDQLIRDGQVTRGFLG 294  
QY 303 VIQEVSYGLAQSGFLDKASGLAKTILPGSPAERAGLOAGDIVLSLDGGIRSSGDLPLV 362  
Db 295 VTIQPIDAELACYKLEKVGALVTDVVKSGPAKAGLRQEDIVVAYNGKEVESLSMFRN 354  
QY 363 MVGAITPGKEVSLGVWRKGEITIKAKLGNAAHTGASSKTDEAPYTEQSGTFSVESAG 422  
Db 355 AVSLMNPOTRIVLKVVRREGKVIETPVTVSQAPKEDGMS-----ALQRVG 398  
QY 423 ITLQTHDSSGKHL-----VVVRVSDAAERAGLRHGDEILAV 459  
Db 399 IRVQNLTPETAKKLGIAETPKGLITISVEPGSVAASGIAPGQILAV 446  
RESULT 7  
HTOA\_HAEIN STANDARD; PRT; 466 AA.  
ID HTOA\_HAEIN  
AC P45129;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable periplasmic serine protease do/hhoA-like precursor  
DE (EC 3.4.21.-).  
GN H11259.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;

	QY	262	SQIYSRSGGPMGISFAIPIDVAMNVAQLKNTKVQRGLGVIIQEVSYGIAOSFGLOKA	321
		:	:   :   :   :   :   :   :   :   :   :   :   :	
	Db	242	TATISPSGGNAGTAFATIPSNOASNLVOOILEFQGVRRGLLIGKGELNADLAKAFNVSAQ	301
		:	:   :   :   :   :   :   :   :   :   :   :   :	
	QY	322	SGALIAKIILPCSPAERAGLAOGDIVLSLDGEIRSSGDLPVMVGCAITTPGKEVSLGVWRKG	381
		:	:   :   :   :   :   :   :   :   :   :   :   :	
	Db	302	QGAFVSEVLPLPKSAEAKGLKAGDITATMNGOKISSFAEIRAKTATTGACGKEISUTLYURDG	361
		:	:   :   :   :   :   :   :   :   :   :   :   :	
	QY	382	EETIKRAKLGNAAEHGTGASSKTDEAPYTEOQSFTFSVESAGITLTQTDSGKHVVVRV	441
		:	:   :   :   :   :   :   :   :   :   :   :   :	
	Db	362	KSHDVKKML-QADSSQLSSKT-ELPALD-----GATLKDYDAKGVKGIETIKI	408
		:	:   :   :   :   :   :   :   :   :   :   :   :	
	QY	442	----SDAERAGLRHGDBELAV	459
		:	:   :   :   :   :   :   :   :   :   :   :   :	
	Db	409	QPNSLAAQR-GLKSGDIIGI	428
		:	:   :   :   :   :   :   :   :   :   :   :   :	
		RESULT	8	
		DEGP_SALT_Y		
	ID	DEGP_SALT_Y	STANDARD;	PRT; 475 AA.
	AC	P26982;		
	DT	01-AUG-1992 (Rel. 23, Created)		
	DT	01-AUG-1992 (Rel. 23, Last sequence update)		
	DT	01-MAR-2002 (Rel. 41, Last annotation update)		
	DE	Protease do precursor (EC 3.4.21.-).		
	GN	DEGP OR HTRA OR PTD OR STM0209.		
	OS	Salmonella typhimurium.		
	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
	OX	Salmonella.		
	NCBI_TaxID=602;			
	[1]	SEQUENCE FROM N.A.		
	RN	STRAIN=C5;		
	RC	MEDLINE=91251770; PubMed=1645840;		
	RX	Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,		
	RA	Ali T., Miller I., Hormaeche C.;		
	RT	"The role of a stress-response protein in Salmonella typhimurium		
	RT	virulence.";		
	RL	Mol. Microbiol. 5:401-407(1991).		
	RN	[2].		
	RN	SEQUENCE FROM N.A.		
	RP	STRAIN=LT2 / SGSC1412 / ATCC 700720;		
	RC	MEDLINE=21534948; PubMed=11677609;		
	RX	McLellan M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,		
	RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,		
	RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,		
	RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,		
	RA	Waterston R., Wilson R.K.;		
	RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium		
	RT	LT2.";		
	RL	Nature 413:852-856(2001).		
	CC	-!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.		
	CC	INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED		
	CC	SPECIFICITY WITH HHOA/DEGO.		
	CC	-!- SUBCELLULAR LOCATION: Periplasmic.		
	CC	-!- INDUCTION: BY HEAT SHOCK.		
	CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE		
	CC	DEGP/DEGO/DEGS FAMILY.		
	CC	-!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.		
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
	CC	the European Bioinformatics Institute. There are no restrictions on its		
	CC	use by non-profit institutions as long as its content is in no way		
	CC	modified and this statement is not removed. Usage by and for commercial		
	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
	CC	or send an email to license@isb-sib.ch).		
	CC	-----		
	EMBL; X54548; CAAB38420.1; -.			
	DR	EMBL; AF008704; AAL1973.1; -.		
	PIR; S15337; S15337.			
	DR	PIR; S21327; S21327.		

DR MEROPS; S01.273; --  
DR StyGene; SG10173; degp.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00595; PDZ; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00834; PROTEASES2C.  
DR SMART; SM00228; PDZ; 2.  
DR PROSITE; PS50106; PDZ; 2.  
DR Hydroxylase; Serine protease; Heat shock; Periplasmic; Signal;  
KW Complete proteome.  
FT SIGNAL 1 26 BY SIMILARITY.  
FT CHAIN 27 475 PROTEASE DO.  
FT DOMAIN 281 372 PDZ 1.  
FT DOMAIN 378 467 PDZ 2.  
FT ACT\_SITE 132 132 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 162 162 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 237 237 CHARGE RELAY SYSTEM (POTENTIAL).  
SQ SEQUENCE 475 AA; 49315 MW; 80685BF3C1A289F CRC64;  
  
Query Match 26.9%; Score 633.5; DB 1; Length 475;  
Best Local Similarity 34.1%; Pred. No. 2.1e-33;  
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;  
  
QY 44 TKDQGSYMLPDPALVQSEPAVV--NIQAAPAPRTQNGSGNAE---TDSPLDSD 97  
Db 29 TSSAMTAQMSLAPLMEKVPVSINVEGSTTVNTPRPNRQFFGGDDSPFCQDGS 88  
QY 98 PFYEFFRLVPMPEIQEEADGG-----LNFSGFII-SKNGVILTNHVVAGM 148  
Db 89 PF-----QNSPFC--QCGNGGNGGQQQKFMALGSGVVIDAAKGVVVTNNHVDNAS 138  
QY 149 SIKVLLNDKRYTAKLIGSDVQSDVALKIDATELPVVKIGNPKNPKLPGEWVAIGAPF 208  
Db 139 VIKVQLSDGRKFDKAVGVKQDPRSDIALIQIONPKNLTAIKLADSDALRGDVTVAIGNPF 198  
QY 209 GFDNSVTAGIVSAKRS-LPNESTYPTFTQDVAINPGNSGGLPFLNKGQVGVINSQIYSR 267  
Db 199 GLGTVTSIGVIALSGSLNLYENFTQDAAINRNGSGGALVNLNGELIGINTAILAP 258  
QY 268 SGGFMGISFAIPIDVAMNVAQLNTKGVORGLGVIIQEVSYGLAQSGFDKASCALIA 327  
Db 259 DGGNIGIGFALPSNWVNLTQMVVEYGVQKRGELGIMTELNSELAKAMKVDAQGFVFS 318  
QY 328 KILPGSPAERAGLQAGDVLSDGGEIRSSGDLPMVWGAITPGKEVSLGVWRKGEIITIK 387  
Db 319 QVMPNSSAARAGIKAGDVITSINGKPISSFAALRAQVCTMPVGSKISLGLLREGKAITVN 378  
QY 388 AKLGNAAEHTGASCKTDEAPYEQSGTFSVESA-----GITLTHTDSSGKHLVVVR 440  
Db 379 LEL-----QSSQSQVDSSTIFSGIEGAEMSNRQDKGVVVSVVK 418  
QY 441 VSDAAERAGLRHGEILAVRASP 463  
Db 419 ANSPAAGIGLKKGDVIIGANQQP 441  
  
RESULT 9  
DEGO\_ECOLI  
ID DEGO\_ECOLI STANDARD; PRT; 455 AA.  
AC P39099;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protease degQ precursor (EC 3.4.21.-).  
GN DEGO OR HHOA OR B3234.  
OS Escherichia coli.  
OC Bacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RA Bass S., Gu O., Goddard A.;  
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.  
RC STRAIN=K12 / W3110;  
RX MEDLINE=96165272; PubMed=8576051;  
RA Waller P.R., Sauer R.T.;  
RT "Characterization of degQ and degS, Escherichia coli genes encoding  
homologs of the DegP protease";  
RL J. Bacteriol. 178:1146-1153(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
CC -!- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
DEGP/DEGO/DEGS FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U15661; AAC3992.1; --  
DR EMBL; U32495; AAC44005.1; --  
DR EMBL; U18997; AAA58036.1; --  
DR EMBL; AE000402; AAC76266.1; --  
DR MEROPS; S01.274; --  
DR SWISS-2DPAGE; P39099; COLI.  
DR ECGene; EGI2612; degQ.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00595; PDZ; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00834; PROTEASES2C.  
DR SMART; SM00228; PDZ; 2.  
DR PROSITE; PS50106; PDZ; 2.  
KW Hydroxylase; Serine protease; Periplasmic; Signal; Complete proteome.  
FT SIGNAL 1 27  
FT CHAIN 28 455 PROTEASE DEGO.  
FT DOMAIN 258 349 PDZ 1.  
FT DOMAIN 355 447 PDZ 2.  
FT ACT\_SITE 109 109 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 139 139 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).  
SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;  
  
Query Match 26.7%; Score 627.5; DB 1; Length 455;  
Best Local Similarity 37.2%; Pred. No. 4.8e-33;  
Matches 155; Conservative 72; Mismatches 141; Indels 49; Gaps 7;  
  
QY 54 LPDFAQLVQSEGPVAVNIQAAPRTQNGSGNAETDSDPLADSDPFYEFFRLVPMPEI 113  
Db 39 LPSLAPMLEKVLPAVVSVRV-----EGTASQGGQKIPBEFKFFGDDLDPDPAQ 86  
QY 114 PQEADDGGLNFGSGFII-SKNGVILTNHVVAGMSKVLNDRKREYTKALIGSDVQSD 172  
Db 87 PFE-----GLSGVVIINASKGVLTNNHVINQAQKISQLNDNGREFDAKLIGSDQSD 139





```
Db 38 MPSLAPMLEKVPMSVSVINV-----EGSTTVNTPRMPRNFQOFFGDDSPFCQEGSPF- 89
QY 108 PNMPEIQEADGG-----LNFSGSFI-SKNGYILTTHVVGMSIKVLNDKREY 160
Db 90 QSSPFCGGGGGGGQQQKFMALGSGVIIDADKGVVVTNNHVVNDATVIKQLSDGRKF 149
QY 161 TAKLIGSDVQSDVALLKIDATEELPVVKIGNPNLKPGEWVAIGAPFGFDSNVTAGIVS 220
Db 150 DAKMVGKDRPSDIALIQINPKNLTAIKMADSDALRVGDYTVTAIGNPFGGCTVTSIGVS 209
QY 221 AKGRS-LPNESTYPTFOTDAVAINPGNSGGPLFNKGOVGVGINSOIYSRSGFMCISFAIP 279
Db 210 ALGRSGLNENYENFTQDAINRNSGGALVNLNGELIGINTAILAPDGGNIGIGFAIP 269
QY 280 IDVMNVAEOLKNTGKVRQGLGVIIQEVSYGLAQSGFLDKASGALIAKIALPSPABRAG 339
Db 270 SNMKVNLTSQWVEYGVQKRGELGIMCTELNSELAKAMKVDAQRGAFFVSQVLPNSSAAKAG 329
QY 340 LQAGDIVLSDGGEIRSSGDLPMVWGCAITPGKEVSLGVWRKGEIITKAKLGNAAHTGA 399
Db 330 IKAGDVITSLNGKPTISSFAALRAQVGTMPVGVSKLTGLLRDQGVNVNLEL----- 380
QY 400 SSKTDEAPYTEQSGGFVESAE-----GITLQTHDSSGKHLVVVVRVSDAAERAGLRH 452
Db 381 -----QSSQNVQDSSSIFNGIEGAEKSGKQGVVNVNVTCTPAAQIGLKK 429
QY 453 GDEIL 457
Db 430 GDVII 434

RESULT 12
DEGP_BUCAP
ID DEGP_BUCAP STANDARD; PRT; 478 AA.
AC P57322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR BU228.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrtosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
MEDLINE=20445173; PubMed=10993077;
Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.:
"Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
Nature 407:81-86(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001118; BAB12943.1; -.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001940; Protease2C.
CC InterPro; IPR001254; Trypsin.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 1.
```

```
KW Hydrolase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 478
FT DOMAIN 116 254
FT DOMAIN 281 372
FT DOMAIN 387 469
FT ACT_SITE 133 133
FT ACT_SITE 163 163
FT ACT_SITE 238 238
SQ SEQUENCE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;

Query Match 25.3%; Score 594.5; DB 1; Length 478;
Best Local Similarity 34.1%; Pred. No. 6.5e-31;
Matches 150; Conservative 86; Mismatches 151; Indels 53; Gaps 10;

QY 43 HTKDDGSVMLLPDPFAQLVQSEGPVAVNIAAPAPRTQNGSNAETDSPL-----A 94
Db 31 NTKNSIVSREISPSLAPMLEKYMPSVISI-----NIEGSAITRTSRLPHQFPFFG 81
QY 95 DSDPF-----YEFKRLVPMPEIPOEADDGGLNF---GSGFIISK-N-CYILTNT 142
Db 82 DNSPFCQGNSPFRHSPFCHNP-----DSDDKKEKFRALGSGVVIINADKGYAVTNH 133
QY 143 VVAGMSIKVLNDKREYTAKLIGSDVQSDVALLKIDATEELPVVKIGNPNLKPGEWVA 202
Db 134 VVENANKIQVQLSDGRRYEARVIGKDSRSDIALIQLKNNANLSEIKIADSDNLRVGDYTV 193
QY 203 AIGAPFGFDSNVTAGIVSAKGRS-LPNESTYPTFOTDAVAINPGNSGGPLFNKGOVGVN 261
Db 194 AIGNPYLGIVTSGIISALGRSGLNIEHYENFIQTDAAINRNSGGALVNLKGEILGIN 253
QY 262 SQIYSRSGGFMGISFAIPIDVAMNVABEOLKNTGKVRQGLGVIIQEVSYGLAQSGFLDKA 321
Db 254 TAILAPDGGNIGIGFAIPCNVKNLTQMVQGVQVRRGELGIMGELNLSDLAQIMKINSQ 313
QY 322 SGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDLPMVWGCAITPGKEVSLGVWRK 381
Db 314 KGAFVSRVLPNSSAFEGAGIKAGDIIISLNRKPIISFSSLRARIEIGSLPVATKMEIGVFRE 373
QY 382 EEITKAKLGNAAHTGASSTDEAPYTEQSGTFSVESAGITLQTHDSSGKHLVV--V 439
Db 374 RIKNITVELKHSVKH-NUNSEND-----YIGIEGVDLSDIYFNEQKVIKVDNV 420
QY 440 RVSDAAERAGLRHGDILAV 459
Db 421 KPHTPASKIGFKKDDIILNV 440

RESULT 13
DEGP_BUCAP
ID DEGP_BUCAP STANDARD; PRT; 478 AA.
AC O85291;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98353428; PubMed=9688822;
RA Thao M.L., Baumann P.;
RT "Sequence analysis of a DNA fragment from Buchnera aphidicola (Aphid
endosymbiont) containing the genes dapD-htra-llvi-llvh-ftsl-ftsi-
murE.";
RL Curr. Microbiol. 37:214-216(1998).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
```



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF060492; AAC32331.1; -  
MEROPS: S01.273; -  
InterPro: IPR001478; PDZ.  
InterPro: IPR001940; Protease2C.  
InterPro: IPR001254; Trypsin.  
Pfam: PF00595; PDZ; 2.  
Pfam: PF00089; trypsin; 1.  
PRINTS: PR00834; PROTEASES2C.  
SMART: SM00228; PDZ; 2.  
PROSITE: PS50106; PDZ; 1.  
Hydrolase: Serine protease; Signal.  
SIGNAL 1 26  
CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.  
DOMAIN 116 254 CATALYTIC.  
DOMAIN 281 372 PDZ 1.  
DOMAIN 387 469 PDZ 2.  
ACT\_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).  
ACT\_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).  
ACT\_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).  
SEQUENCE 478 AA; 51303 MW; C04824F7EF4E98E CRC64;

Query Match 24.38; Score 571.5; DB 1; Length 478;  
Best Local Similarity 32.48; Pred. No. 1.9e-29;  
Matches 142; Conservative 81; Mismatches 146; Indels 69; Gaps 9;

Qy 53 LLPOFAVLQSEGPAV--VNIOAAPAPRTQN-----GSGNAETDSDPLAD 95  
Db 41 LAPSLAPMLEKVMFSVINETGTVVTSRLPHQFPFGHNSPFCOGNSPFRSPCR 100  
Qy 96 SDP-----PYEPRKRLVPMPEIPOEADDDGGLNFGSGFIISKN-GYILTNTHVWAGMSI 150  
Db 101 SNPNSSMHEKPHAL-----GSGVIINADKAYAVTNHNVENANKI 141  
Qy 151 KVLNDRKREYAKLIGSDVQSDVALLKIDATEELPVVIGNPKNLKPEWAAICAPGFG 210  
Db 142 QVQLSDGRYFASIGKDSRDIALIQLKNAKLSAIKAIDSDTLRVGDYTVVAINGPYGL 201  
Qy 211 DNSVTAGTVSAKGRS-LPNESVTPFIQTDVAIIPGNSGGPLNKGQVGVGINSQIYSRG 269  
Db 202 GETVTSGLIISALGRSLNIHYENFIQTDVAIIPGNSGGALVNLKGLIGINTAILAPDG 261  
Qy 270 GFMGISFAIPIDVAMNVAEQLKNTGKVGQGLGVIIQEVSYGLAOSFGLDKASGALIAKI 329  
Db 262 CNIIGFAIPGNMVKNLTEQMKVQGVKRGELGIIGMELNSDLAHVMKINAKQKGAFFSQV 321  
Qy 330 LPGSPAERAGLQAGDIVLSLGGGEIRSSGDLPLVVMGATTPGKEVSLGVWRGEE----IT 385  
Db 332 LPNSAFHAGKAGDIIVSLNKKTISSFAALRAEVGSLPVSTKMEIGFRNGITKNVIVE 381  
Qy 386 IKAKLGNAAE----HTGASSKTDEAPYTEQSGTFSVESAGITLTQHTDSSGKHLVVVVR 441  
Db 382 LKPSLKNSVSLGDIYTG-----IEGADLSDCSLNGQKGVKIENTIKL 422  
Qy 442 SDAERAGLRHGDITLAV 459  
Db 423 NTOASKIGFKKDDIIVEV 440

RESULT 14

DEGI\_ARATH  
ID DEGI\_ARATH STANDARD; PRT; 437 AA.  
AC O22609; O9LK85;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)  
Protease DO-like 1, chloroplast precursor (EC 3.4.21.-).  
DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
CC NCBI\_TaxID=3702;  
[1]  
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RP MEDLINE=98175982; PubMed=9507020;  
RX Itzhaki H., Naveh L., Lindahl M., Cook M., Adam 2.;  
RA "Identification and characterization of DegP, a serine protease  
RT associated with the luminal side of the thylakoid membrane.";  
RL J. Biol. Chem. 273:7094-7098(1998).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CV. COLUMBIA;  
RX MEDLINE=20363099; PubMed=10907853;  
RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the 4,251,695 bp regions covered by 90 Pl, TAC  
RT and BAC clones.";  
RN DNA Res. 7:217-221(2000).  
[3]  
RN SEQUENCE OF 104-118.  
RP STRAIN=CV. COLUMBIA;  
RX Kieselbach T., Bystedt M., Schroeder W.P.;  
RA Submitted (JUL-2000) to the SWISS-PROT data bank.  
RL  
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.  
CC MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,  
CC CAN DEGRADE BETA-CASEIN.  
CC -!- ENZYME REGULATION: INHIBITED BY PHENYL METHYL SULFONYL FLUORIDE AND  
CC O-PHENANTHROLINE.  
CC -!- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID  
CC MEMBRANE.  
CC -!- INDUCTION: BY HEAT SHOCK.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
CC DEGP/DEGO/DEGS FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF028842; AAC39436.1; -  
EMBL: AP000371; BAB02539.1; -  
EMBL: AP001302; BAB02539.1; JOINED.  
MEROPS: S01.279; -  
Mendel: 25267; Arath: 3069; 25267.  
InterPro: IPR001478; PDZ.  
InterPro: IPR001940; Protease2C.  
InterPro: IPR001254; Trypsin.  
Pfam: PF00595; PDZ; 1.  
Pfam: PF00089; trypsin; 1.  
PRINTS: PR00834; PROTEASES2C.  
SMART: SM00228; PDZ; 1.  
PROSITE: PS50106; PDZ; 1.  
Hydrolase: Serine protease;  
TRANSIT 1 ?  
CHAIN 104 437 PROTEASE DO-LIKE 1.  
DOMAIN 152 321 SERINE PROTEASE.  
DOMAIN 324 421 PDZ.  
ACT\_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).  
ACT\_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).  
ACT\_SITE 280 280 CHARGE RELAY SYSTEM (POTENTIAL).  
CONFLICT 12 23 HSPSSLNSNT -> SSTFLHSPSSHL (IN REF. 2).



Db 201 : |||||:|:| ||| :|| | : | | : | :|:|:| 255  
Qy 178 IDATEELPVVKIGNPKNLKPGEWAAIGAPFGFNDNSVTAGIVSAK---GRSLP-NESYTP 233  
Db 256 IDHOGKLPVLLGRSSSELRPGFVVAIGSPSLQNTVTGIVSTTORGKELGLRNSDMD 315  
Qy 234 FIOTDVAINPNSGGPLFNLKGQVVGINSQIYRSRGGFMGISFAIPID-----VAMNVAE 288  
Db 316 YIOTDAIINYNSGGPLVNLDEGIVGINTLKVT-----AGISFAIPSDKIKKFLTESHDR 370  
Qy 289 QLKNTGKVORQOLGVIIQEVSYGLAQSEF-----LDKASGALIAKILPGSPAERAGLOA 342  
Db 371 QAKGKAITKKYIGIRMMSLTSSRAKELKDRHRDPDVISGAYIIIEVIPDTPAEAGGLKE 430  
Qy 343 GDIVLSLDGGEIRSSGDLPMVMGATPGKEVSLG-VWRKGE---ITI 386  
Db 431 NDVIISINGQSVSANDVSDVI-----KRESTLNNVVRGNEIMITV 473

Search completed: June 10, 2002, 12:14:14  
Job time: 223 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:56 ; Search time 31.16 Seconds  
(without alignments)  
2581.600 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFYKQYFALACALLAG.....ERAGLRHGDELLAVRASPRQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL19:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvrius:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_cheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2246	95.5	499	16 Q9JVT1	Q9jvt1 neisseria m
2	989.5	42.1	474	2 Q9ALS1	Q9als1 pseudomonas
3	985.5	41.9	474	16 Q57155	Q57155 pseudomonas
4	894.5	38.0	481	2 Q9AOD1	Q9aod1 pseudomonas
5	862.5	36.7	514	16 Q9PBA3	Q9pba3 xylella fas
6	746	31.7	513	16 Q9BCS8	Q9bcs8 rhizobium l
7	743	31.6	473	2 Q44476	Q44476 azotobacter
8	719.5	30.6	516	16 Q985F9	Q985f9 rhizobium l
9	673	28.6	504	16 Q98KJ1	Q98kj1 rhizobium l
10	664.5	28.2	500	2 Q44652	Q44652 brucella ab
11	660	28.0	433	2 Q9FDL1	Q9fdl1 aeromonas h
12	639.5	27.2	478	2 P74978	P74978 yersinia en
13	638	27.1	511	16 Q92JA1	Q92ja1 rickettsia
14	637	27.1	437	2 Q68197	Q68197 haemophilus
15	637	27.1	459	16 Q9CMS7	Q9cms7 pasteurella
16	637	27.1	474	2 Q44596	Q44596 brucella ab

17	636	27.0	503	16 Q926C8	Q926c8 rhizobium m
18	636	27.0	530	16 Q9A4S2	Q9a4s2 caulobacter
19	632.5	26.9	456	16 Q9KUF5	Q9kuf5 vibrio chol
20	632	26.9	463	2 Q68198	Q68198 haemophilus
21	621.5	26.4	500	2 Q9KJN6	Q9kjm6 myxococcus
22	614.5	26.1	459	16 Q9WZ41	Q9wz41 thermotoga
23	606.5	25.8	452	16 P73354	P73354 synecocyst
24	606.5	25.8	491	2 Q9L8K0	Q9l8k0 shigella so
25	604.5	25.7	481	16 Q9PGL3	Q9pgl3 xylella fas
26	603.5	25.6	371	2 Q31388	Q31388 bradyrhizob
27	600.5	25.5	416	16 P73940	P73940 synecocyst
28	600	25.5	460	2 Q06439	Q06439 rhodobacter
29	589	25.0	389	16 Q9HVL1	Q9hvl1 pseudomonas
30	582	24.7	476	16 Q9ZM18	Q9zm18 helicobacte
31	578	24.6	428	16 Q98N31	Q98n31 rhizobium l
32	575	24.4	464	2 Q56885	Q56885 yersinia en
33	574.5	24.4	443	16 Q25663	Q25663 helicobacte
34	571	24.3	453	16 Q67436	Q67436 aquifex ao
35	545.5	23.2	394	16 P72780	P72780 synecocyst
36	545	23.2	472	16 Q9A8R9	Q9a8r9 caulobacter
37	544.5	23.1	465	16 Q92QE6	Q92qe6 rhizobium m
38	529	22.5	466	2 Q53247	Q53247 rickettsia
39	524	22.3	466	2 Q05335	Q05335 rickettsia
40	524	22.3	466	2 Q53246	Q53246 rickettsia
41	522	22.2	466	2 Q53249	Q53249 rickettsia
42	511.5	21.7	472	16 Q9PN69	Q9pn69 campylobact
43	511	21.7	466	2 Q53251	Q53251 rickettsia
44	509	21.6	389	2 Q51374	Q51374 pseudomonas
45	504.5	21.4	472	2 Q46120	Q46120 campylobact

ALIGNMENTS

RESULT	ID	Q9JVT1	PRELIMINARY;	PRT;	499 AA.
AC	Q9JVT1	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	PUTATIVE PERIPLASMIC SERINE PROTEASE (EC 3.4.21.)				
GN	NMA0710				
OS	Neisseria meningitidis (serogroup A)				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria				
OX	NCBI_TaxID=65699;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;				
RX	MEDLINE=20222556; PubMed=10761919;				
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S., Jagals K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;				
RT	"Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."				
RL	Nature 404:502-506(2000).				
CC	- I - SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.				
DR	EMBL: AL162754; CAB83996.1; -				
DR	MEROPS: S01.UFC; -				
DR	InterPro: IPR001478; PDZ.				
DR	InterPro: IPR001940; Protease2C.				
DR	InterPro: IPR001254; Trypsin.				
DR	Pfam: PF00595; PDZ; 2				
DR	Pfam: PF00089; trypsin; 1.				
DR	PRINTS: PR00834; PROTEASES2C.				
DR	SMART: SM00228; PDZ; 2.				
DR	PROSITE: PS0106; PDZ; 1.				
DR	PROSITE: PS0240; TRYPSIN_DOM; 1.				
KW	Complete proteome; Hydrolase; Protease; Serine protease.				
SEQUENCE	499 AA; 52563 MW; 73E979F0A7F5B969 CRC64;				

```
Query Match          95.5%; Score 2246; DB 16; Length 499;
Best Local Similarity 95.9%; Pred. No. 6.1e-132;
Matches 444; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 VFKKYQYFALAALCAALLACERAGSFFGADKKKEASVERIEHTKDDGSVSMLLPDFQAQL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MFKKYQYLAALAALCAASLAGCDKAGSFFGADKKKEASVERIEHTKDDGSVSMLLPDFVQL 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 VOSEGPVAVNIQAAPRTONGSGNAETSDPLADSDPFYEFFKRLVNPMPPEIQEADD 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VOSEGPVAVNIQAAPRTONGSSNAETSDPLADSDPFYEFFKRLVNPMPPEIQEADD 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 GGLNFGSGFIISKNGYILTNTHVAGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GGLNFGSGFIISKNGYILTNTHVAGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 TEPLPVKIGNPNKLPGEWVAAGAPFGFDSVNTAGIVSAKGRSLPNESTPTPTQTDVA 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 TEPLPVKIGNPNKLPGEWVAAGAPFGFDSVNTAGIVSAKGRSLPNESTPTPTQTDVA 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 241 INPNSGGPLFNKLGQVVGINSQIYSRSGFMGISFAIPIDVAMNVAEOLKNTCKVORGQ 300
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 INPNSGGPLFNKLGQVVGINSQIYSRSGFMGISFAIPIDVAMNVAEOLKNTCKVORGQ 300
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 301 LGVIIQEVSYGLAQSEGLDKAGGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 LGVIIQEVSYGLAQSEGLDKAGGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 361 PVMVGAITPGKEVSLGWWRKGEITITKALGNAAEHTGASKTDEAPYTEQOSGTFSVES 420
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 PVMVGAITPGKEVSLGWWRKGEITITKALGNAAEHTGASKTDEAPYTEQOSGTFSVES 420
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 421 AGITLTHTDSSGKHLVVVRSVDAERAGLRHGDGDEILAVRAS 463
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 AGITLTHTDSSGKHLVVVRSVDAERAGLRHGDGDEILAVRAS 463
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
Q9ALS1 PRELIMINARY; PRT; 474 AA.
AC Q9ALS1
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCD.
DE MUCD.
OC Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCBPP-PA14;
RA Yorgey P.S., Rahme L.G., Tan M., Ausubel F.M.;
RT "The Roles of mucD and Alginate in the Virulence of Pseudomonas
RL aeruginosa in Plants, Nematodes, and Mice.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF343973; AAK11276.1;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 474 AA; 50349 MW; A22PD4338B859D4C CRC64;
```

```
Query Match          42.1%; Score 989.5; DB 2; Length 474;
Best Local Similarity 52.5%; Pred. No. 9.1e-54;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

QY 54 LPDFAQLVQSEGPVAVNI---QAAPAPRTQNGSGNAETSDPLADSDP-FYEFFKRLVNP 109
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 27 LPDFTPLVEQASPAVAVNISTROKLP-----DRAMARQQLSIPDLEGUPPMFRDFLERSIPQ 82
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 110 MFEIP---QEEADGGNFGSGFIISKNGYILTNTHVAGMSIKVLLNDKREYTAKLIG 166
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 83 VPRNPRGQOREAQ-----SLGSGFIISNDGYILTNHVVADAEILVRLSDRSEHKAKLVG 138
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 167 SDVQSDVALLKIDATEELPVKIGNPNKLPGEWVAAGAPFGFDSVNTAGIVSAKGRSL 226
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 ADPRSDVAVLKTEA-KNLPTLKLGDSNKLVGEWLAIGSPFGFDHSTAGIVSAKGRSL 197
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 227 PNESTPTPTQTDVAINPNSGGPLFNKLGQVVGINSQIYSRSGFMGISFAIPIDVAMNV 286
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 PNESTVPFPTQTDVAINPNSGGPLLNLEGEVVGINSQIFTRSGGPMGLSFAIPIDVALNV 257
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 287 AEOLKNTGKVORGLQCVIIQEVSYGLAQSEGLDKAGGALIAKILPGSPAERAGLQAGDIV 346
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 258 ADOLKAKGVSRGVLGVQEVNKNOLAESFGLDKPSGALVAQLVEDGPAKGGLOVGDVI 317
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 347 LSLDGEIRSSGDLPMVVGAITPGKEVSLGWWRKGEITITKALGNAAEHTGASKTDE- 405
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 318 LSLNGQSINESADLPHLVGNMKPGDKINLDVIRNGQRKSLSMVAGNLPD-----DDEE 370
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 406 ----APYTEQOSGTFSVESAGITLTQTH--TDSSGKHLVVVRSVDA-AERAGLRHGDG 456
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 371 IASMGAPGAERSNRLGVTVADLTAEQRKSLDIQG-GVVIKEVODGPAAVIGLRPGDVI 428
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
Q57155 PRELIMINARY; PRT; 474 AA.
AC Q57155.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCD (SERINE PROTEASE MUCD).
DE MUCD OR PA0766.
CN Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=93391358; PubMed=8378309;
RA Martin D.W., Schurr M.J., Mudd M.H., Govan J.R., Holloway B.W.,
RA Deretic V.;
RT "Mechanism of conversion to mucoidy in Pseudomonas aeruginosa
RT infecting cystic fibrosis patients.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=95286510; PubMed=7768826;
RA Yu H., Schurr M.J., Deretic V.;
RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas
RT aeruginosa AlgU: E. coli rpoE restores mucoidy and reduces sensitivity
RT to reactive oxygen intermediates in algU mutants of P. aeruginosa.";
RL J. Bacteriol. 177:3259-3268(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=96134987; PubMed=8550474;
RA Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd M.H., Yu H.,
RA Deretic V.;
RT "Two distinct loci affecting conversion to mucoidy in Pseudomonas
RT aeruginosa in cystic fibrosis encode homologs of the serine protease
RT HtrA.";
```

```
RL J. Bacteriol. 178:511-523(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; U49151; AAC43718.1; -.
DR EMBL; U32853; AAC43676.1; -.
DR EMBL; AE004511; AAG04155.1; -.
DR MEROPS; S01.UFC; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Complete proteome; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 474 AA; 50321 MW; 8AB5D2A89867BEBE CRC64;

Query Match 41.9%; Score 985.5; DB 16; Length 474;
Best Local Similarity 52.5%; Pred. No. 1.6e-53;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

QY 54 LPDFAQLVQSEGPVAVNI---QAAPAPRTQNGSNAETSDPLADSDP--FYEFFKRLVP 109
DB 27 LPDFTPLVEQASPAVVNISTRQKLP-----DRAMARGQLSDPLDLEGLPPMFRDLESLTPQ 82
QY 110 MPPIP---QSEADDDGLNFGSGFIISKNGYILTTTHVVAGMGSIKVLNDRKREYTA 166
DB 83 VPRNPRCQREQA---SLSGGFIISNDGYILTTNNHVVADADEILVRLSDRSEHKAKLIG 138
QY 167 SDVQSDVALLKIDATEELPVVKNPKNKPGEVAAIGAPFGFDNSVTAGIVSAKGRSL 226
DB 139 ADPRSDVAVLKIEA-KNLPTLKLGDNSNKLKVGWVLAIGSPFGFDHVSVTAGIVSAKGRSL 197
QY 227 PNEYTPFTIOTDVAINPNSGGPLNFKGVVGINQSIYRSRSGFGMGISFAIPIDVAMNV 286
DB 198 PNEYVPFTIOTDVAINPNSGGPLNQLQGEVVGINSQIFTRSGFGMGISFAIPIDVALLNV 257
QY 287 AEOLKNTGKVQRCOLGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346
DB 258 ADQLKAGKGVSRGWLGVVIEVKNDLAESEGLDKRPSGALVAQLVEDGPAAGGLQGVGDI 317
QY 347 LSLDGGIEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEETITKAKLGNAAEHTGASKTDE- 405
DB 318 LSLNGOSINESADLPHLVGNMKGDKINLDVIRNGQRKSLSMVAGSLPD-----DDEE 370
QY 406 -----APYTEQQSGTFSVESAGITLQTH--TDSSGKHLVVVVRSDA-AERAGLRHGEI 456
DB 371 IASMGAPGAERSSNRGLGTVTADLTAEQKKSLLDIQG-GVVIKEVQDGPAAVIGLRPGDVI 428

RESULT 4
Q9AQD1 PRELIMINARY; PRT; 481 AA.
AC Q9AQD1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCL.

GN MUCL.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FF5;
RA Keith L.M.W., Bender C.L.;
RT "Genetic divergence in the algT-muc operon controlling alginate
RT biosynthesis and response to environmental stress in Pseudomonas
RT syringae.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190580; AAK01318.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 481 AA; 50601 MW; E2C9C11137B83920 CRC64;

Query Match 38.0%; Score 894.5; DB 2; Length 481;
Best Local Similarity 47.6%; Pred. No. 7.6e-48;
Matches 202; Conservative 55; Mismatches 128; Indels 39; Gaps 8;

QY 54 LPDFAQLVQSEGPVAVNI---QAAPAPRTQNGSNAETSDPLADSDP--FYEFFKRLVP 108
DB 30 LPDFTPLVEQASPAVVNISTRQKLPDRAVAN-----QOMPDLGLEPLPMLREFLESLMP 82
QY 109 NMPEIPQEEADDDG-----GLNFGSGFIISKNGYILTTTHVVAGMGSIKVLNDRKREYTA 163
DB 83 PGRSPPGSGAGKGDROREAOQSLGSGFIISPDGYVLTNNHVIDGADELVLRLSDRSELKAK 142
QY 164 LIGSDVQSDVALLKIDATEELPVVKNPKNKPGEVAAIGAPFGFDNSVTAGIVSAK 223
DB 143 LVGTDPRTDVAVLKIEG-KDLPTAKLGNSNTLVKGEVWVLAIGSPFGFDHVSVTAGIVSAK 201
QY 224 RSLPNESYTPFTIOTDVAINPNSGGPLNFKGVVGINQSIYRSRSGFGMGISFAIPIDVA 283
DB 202 RSLPNDDYVPFTIOTDVAINPNSGGPLNFMAGVVGINSQIFTRSGFGMGISFAIPIDVA 261
QY 284 MNVAEOLKNTGKVQRCOLGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAG 343
DB 262 MDVANOLKANGKVSRLGWLGVVIEVKNDLAESEGLDKPAGALVAQLVEDGPAAGKGVQVG 321
QY 344 DIVLSLDGGIEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEETITKAKLGNAAEH----- 396
DB 322 DVILSANGQPIVMSADLPHLIGNLKDGSKAELEVIDRGKRLKTLVTVGALPDEQEGMDV 381
QY 397 --TGASSKTDEAPYTEQQSGTFSVESAGITLQTHDSSGKHLVVVR--VSDAERAGLRH 452
DB 382 GGTGA-----ERSSNRGLGVSVELTAEQKKSLLDKGKGVAIKEVTGPGASLIGLQA 431
QY 453 GDEI 456
DB 432 GDVI 435

RESULT 5
Q9PBA3 PRELIMINARY; PRT; 514 AA.
AC Q9PBA3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PERIPLASMIC PROTEASE.
GN XF2241.
```

```
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.S.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro L.R.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramie E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vagla H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RA "The genome sequence of the plant pathogen Xylella fastidiosa.";
RT Nature 406:151-159(2000).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AF04037; AF85040.1; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR000126; Ser_proteas-V8.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR Complete proteome; Hydrolase; Serine protease.
SQ SEQUENCE 514 AA; 54140 MW; 707C23FD3C82BE4C CRC64;
```

```
Query Match 36.7%; Score 862.5; DB 16; Length 514;
Best Local Similarity 46.2%; Pred. No. 8.2e-46;
Matches 194; Conservative 64; Mismatches 129; Indels 33; Gaps 10;
```

```
Qy 54 LPDFAOLVSEGPVAVVNIQAAPRTQNGSGNAETSDPLADSPFFYEKKRL-----V 107
Db 54 LPDFTQLVDQVGPGVNIETVIRKKYVGRKRGRIPLND-----IPEFFRRFGPDQFM 106
Qy 108 PNMPPEIQEADDGGL--NFGSGFIISKNGYILTNTHVAVMGSIKVLNDRKREYAKL 164
Db 107 PNQPRGQD--DEGGTAGRMGSGFIISKDGYILTNHHVITGASEVTIKLTDREFRAKI 164
Qy 165 IGSVDQSDVALLKIDATEELPVYKIGNPKNKPGEVAAIAGPFGDNSVTAGIVSAKR 224
Db 165 IGSDEQYDVALLKIDA-KNLPTVRIGDSSSLKSGQWVAIGSPGLDHSVTAGIVSALGR 223
Qy 225 SLPNES-YTPFTQTDVAINPGNSGGFLNLKQGVVINSQIYRSKSGFGMGISFAIPDVA 283
Db 224 STSDQRYVPFIQTDPVPIQNGNSGGFLNTRGEVIGINSQIFSAASGCGMGISFAIPINLA 283
```

```
Qy 284 MNVAEOLKNTGKVQRGOLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAG 343
Db 284 INAAEIQRTKGTQVRSMLGVEIGPIDALKAQGLGLPDSRCALYNNIPPHSPAAGKIEVG 343
Qy 344 DIVLSLDGGEIRSSGDLPMVGAITPGKEVSLGWKRG--EITIKAKLGNAAHTGASS 401
Db 344 DVIRSVNGKVISFSFDLPLIGMMPCTKATGLIIRDGKPREIVVTLSALN-ODATASED 402
Qy 402 KTDPEIYTEQOQGTFSVESAGITLQTHDSSGKHLVVVRVSDAAERAGLRHGDILAVRA 461
Db 403 EDDTAPSKPETA--NVELLGLQVENLSAARERL-----ASSONAKGGVRITAVTA 452
RESULT 6
Q98CS8 PRELIMINARY; PRT; 513 AA.
ID Q98CS8;
AC Q98CS8;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SERINE PROTEASE, HTRA/DEQ/DEGS FAMILY.
GN MLI5022.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003005; BAB51543.1; -.
DR MEROPS: S01.UPC; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR Complete proteome.
KW Protease; Complete proteome.
SQ SEQUENCE 513 AA; 52258 MW; 901PAF72004F5E3B CRC64;
```

```
Query Match 31.7%; Score 746; DB 16; Length 513;
Best Local Similarity 39.0%; Pred. No. 1.5e-38;
Matches 184; Conservative 81; Mismatches 177; Indels 30; Gaps 11;
```

```
Qy 10 LAALCAALLAGCEKAGSFFGADKKEASFYVERIEHTKDDGSVMLLPD--FAQLVQSEGA 67
Db 12 VAALLGAALLIIPVVVSF--AOSANSTGYSKIVATTQTPVAGITAPNGSFAPVAAADKPA 69
Qy 68 VVNIQAAAPRTQNGSGNAETSDPLADSPDPEYEFKRLVPMN---PEIQEADDGGL 123
Db 70 VVTVTS-----IMKQAPASDDDDGMPLGNS-PFDQYFRQFGDGMGMAPOTPPQAAAREA 123
Qy 124 NFGSGFIISKNGYILTNTHVAVMGSIKVLNDRKREYAKLIGSDVQSDVALLKIDATEE 183
Db 124 -LGSGFIVTADGTVTNHNVVDGASSIKVTLDDGTLPKLVGRDAKNDLAVLKIKSDKP 182
Qy 184 LPVYKIGNPKNKPGEVAAIAGPFGDNSVTAGIVSAKGRSLPNESYTPFIQTDAINP 243
Db 183 LPTVKWGSDDLMTGDQVLAIQNPFGITGTVTAGIVSARGDLHSGPFDFFIOTDAPINH 242
Qy 244 GNSGGFLNLKQGVVINSQIYRSKSGFGMGISFAIPIDVAMNVAEOLKNTGKVQRGOLCV 303
```



```
Db 243 GNSGGPLVDVNGVNGVINTAIYSPNGSGVGVGAIPSDQAQVVKLMKDGSIQYGLGV 302
QY 304 IIEVSYGLAQSGFLDKASGLAKILPGSPAERAGLQAGDVLVSLDGGIRSGDLPVM 363
Db 303 EIQEVPDVASAIGLDHAGALSKVNDSPASASAGVEAGDVITGAFQDVDPKDLRA 362
QY 364 VGAITPKVSLGVWRKGEITTKAKLG-NAAEHTGASSKTDAPYTEQSGTFSVESAG 422
Db 363 VADVAPGAKETLDVWRKGAQISVEVQNSDDVKTASGESAPSAEQGS---RAPAIG 419
QY 423 ITLQTHD-----SSGKH-LVVVRVS--DAERAGLRHGDDEILAVRASP 463
Db 420 LGLMDITPDRQENLNAGNEHGAIVARNPDKAAAGAIQPGDIIVAVNQAP 471
```

```
RESULT 7
ID Q44476 PRELIMINARY; PRT; 473 AA.
AC Q44476;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCD.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Azotobacter.
OX NCBI_TaxID=354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW 136;
RX MEDLINE=96178940; PubMed=8606151;
RA Martinez-Salazar J.M., Moreno S., Najera R., Boucher J.C., Espin G.,
RA Soberon-Chavez G., Deretic V.;
RT "Characterization of the genes coding for the putative sigma factor
RT Algu and its regulators MucA, MucB, MucC, and MucD in Azotobacter
RT vinelandii and evaluation of their roles in alginate biosynthesis.";
RL J. Bacteriol. 178:1800-1808(1996).
EMBL: U30799; AAB01513.1; .
DR MEROPS: S01.UPC; .
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 2.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 473 AA; 50282 MW; C97B357D897738AB CRC64;
```

Query Match 31.6%; Score 743; DB 2; Length 473;  
Best Local Similarity 42.8%; Pred. No. 2e-38;  
Matches 182; Conservative 58; Mismatches 159; Indels 26; Gaps 8;

```
QY 50 VSMLLPDAQLVQSGPAAVNI---QAAPRTONGSGNAETSDPLADSDPFYEFKRL 106
Db 25 VQALPEFTSLVEASPAVNISTRQKLPDRSTVQGLPDL-GLPPL-----PREFLERS 78
QY 107 VPNNPEIQEADGGGLNGFGSGFTISKNGVILTNTHVVAGMGSTKVLNDKREYTKALIG 166
Db 79 IPQLPRTPDNGRREAHSLGSGFTIISPDGVLTNNHVVADADELIVRLSRSELEAEVLG 138
QY 167 SDVQSDVALLKIDATEELPVVYKIGNPNKLPGEVAAIAGPFGDNSVTAGIVSAKGRSL 226
Db 139 ADPLTRCSFVEGQGFES-PHSQTGTYPRIESRGMGSHRFPFRFRSPRDCGHCHCAKP 197
QY 227 PNEYSTPFIQTDVAIAPNCGSGGLFLNLKGVVINSIYSRSGFMGISFAIPIDVAMNV 286
Db 198 AERELVPFIQTDVAIAPNCGSGGLFLDLGRVIGINSQIFTRSGFMGLSFAIPIEVAMGV 257
QY 287 AEOLKNTGKVORGOLGVIIQEVSYGLAQSGFLDKASGLAKILPGSPAERAGLQAGDIV 346
Db 100 LKNTGKVORGOLGVIIQEVSYGLAQSGFLDKASGLAKILPGSPAERAGLQAGDIV 346
```

```
Db 258 ADQLKATGKVGWGLVYIIQEVNKNDLAESFGLDRPAGALVAQVLEDDGPADGGLOVGDVI 317
QY 347 LSLDGGIRSGDLPVMVWGAIITPGKEVSLGVWRKGE-----ITIKA--KLGNAAEHTGAS 400
Db 318 LSLDGHFIVMSADPLHLVGLKPGAAANLEVVVRDGRNRNIAITVGCALPEENGQVQPSIAG 377
QY 401 SKTDEABYTTQQSGTFSVESAGITLQTHDSSGKHLVVVR--VSDAERAGLRHGDDEILA 458
Db 378 -----TEQSSNRLGVTVTTELTAEQKSLDLKGGVIREVLNPGPALIGLRPGDVVTH 429
QY 459 VRASP 463
Db 430 LNNOP 434

RESULT 8
ID Q985F9 PRELIMINARY; PRT; 516 AA.
AC Q985F9;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SERINE PROTEASE.
DE MLR7692.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
EMBL: AP003012; BAB54103.1; .
DR MEROPS: S01.273; .
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR000126; Ser_proteas_V8.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
KW Protease; Complete proteome.
SQ SEQUENCE 516 AA; 53704 MW; FE8DAADC099ABDC6 CRC64;
```

Query Match 30.6%; Score 719.5; DB 16; Length 516;  
Best Local Similarity 36.2%; Pred. No. 6.6e-37;  
Matches 174; Conservative 80; Mismatches 177; Indels 49; Gaps 9;

```
QY 10 LAALCAALLACCEKAGSFFFGADKKEASFVERIEHTKDDGSVMLLPDAQLVQSGPAVV 69
Db 16 LAAASVAVAGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIGVIG 67
QY 70 N--IOAAPAPRTONGSGNAETSDPLADSDPFYEFKRLVPMPEIQEADGGGLN--- 124
Db 68 SVKVKAIQIPTADGSD-ODGFNLNPNQLRFRFFKEF-----RGFGDQGGQNDG 118
QY 125 -----FGSGFTISKNGVILTNTHVVAGMGSTKVLNDKREYTKALIG 166
Db 119 HRRFGHGRDRSNDQPRPVAQSGGFFISEDGLVLTNNHVVEEGTAFVTVNDGKELDAKLVG 178
QY 167 SDVQSDVALLKIDATEELPVVYKIGNPNKLPGEVAAIAGPFGDNSVTAGIVSAKGRSL 226
```

```
Db 179 TDPRDVLAVKGGGKFTYVDEADDSKVRVGDWVAVGNPFGLGGTVTAGIVSARGRDI 238
Qy 227 PNESTYPTFIOTDVAINPNSGGPLNFKGOVVGINSOISYRSGFMGIFSFAIPIDVAMNV 286
Db 239 GAGPYDDFDGLDASVNGSGGPTFNQVVGINTAIFSPSGSVGIAFDIPASTAKQV 298
Qy 287 AEQKNTKGVQVQVGLVYIIEVSYGLAQSFGLDKASGALIAKILPSPFAERAGLQAGDIV 346
Db 299 VEDLMKSGAVQVGLVYIIEVSYGLAQSFGLDKASGALIAKILPSPFAERAGLQAGDIV 358
Qy 347 LSLDGGEIRSGDPLVMVGAITPGKVSGLVWRKGBEITIKAKLGN-AAETHTGASSKTDE 405
Db 359 TVQEGKDVASPKELARLIGAYSPKSVVDVTVWRDCKSQTIKVDGLKLPASDKQASNDQOQ 418
Qy 406 APVTEQSGSFVESAGITLQTHDSSGKHLVVRV--SDAERAGLRHGDILAVRAS 462
Db 419 OPAAPAKPDTL----ADLGLTVTKSENGKGLVVDVDPESAAADR-GIOPGDIITAVNSN 473
RESULT 9
ID Q98KJ1 PRELIMINARY; PRT; 504 AA.
AC Q98KJ1;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PROBABLE SERINE PROTEASE.
GN MLL1451.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RC STRAIN=NAFE303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002997; BAB48823.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR003975; Shal_channel.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR01497; SHALCHANNEL.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; Complete proteome.
KW SEQUENCE 504 AA; 52128 MW; 703EAF6C370AB38A CRC64;
```

Query Match 28.6%; Score 673; DB 16; Length 504;  
Best Local Similarity 36.4%; Pred. No. 5e-34;  
Matches 178; Conservative 73; Mismatches 170; Indels 68; Gaps 15;

```
Qy 10 LAALCAALLAGCEKAGSF----FGADKKEASFVERIEHTKDDGSGVSMLLPDPFAQLVQSEG 65
Db 14 LIAGTAALLVGTVAAPSFVTPVFAA-----DGPASV-----ADLAQGV 52
Qy 66 PAVNITQAAPARTQNGSGNAETSDSPDLPADSPFEPFKRLVPMNPETPOEADGG--- 122
Db 53 GAVNITSTQTVKGTGPG--AVPMQLPGSPFQDFEDDFKNR----GGDKDNSGQV 106
```

```
Qy 123 LNFSGSFTI-SKNGYILTNTHVVAGMSIKVLNDKREYTKAGLSQVSDVALLKIDAT 181
Db 107 QSLGSGFVIDAEQGIIVVTNNHVIADADDIEVNFSDGVTLKATLVGTDTKTDAVLKVDPK 166
Qy 182 -BELPVVTKGNPKNLKPGEWAAIGAPPDFDNSVTAGIVSAKGRSLPNSYTFPIQTDVA 240
Db 167 GHKLTAVKFGDSTKMRVGDWVAVGNPFGLGGTVTVGIVSARNRDNISGYPYDDFIQTDAA 226
Qy 241 INPGSGGPLENLKGOVVGINSOISYRSGFMGIFSFAIPIDVAMNVAEQLNKTKGVQVGO 300
Db 227 INRGSGGPLFNAGVEIGINTAIIISPGSGSIGIFSPSOLASGVVDVQVQVGETRRGW 286
Qy 301 LGVITQEVSYGLAQSFGLDKASGALIAKILPSPFAERAGLQAGDIVLSLDGGEIRSGDL 360
Db 287 LGVRIQVPTDDIAESLGMATAKALVAGVIKGGPVDNGTIQAGDVIIKFGDKDIHEMRDL 346
Qy 361 PMVGAITPGKVSGLVWRKGBEITIKAKLGNAAE-----HTGASSKTDE-----APYEQ 411
Db 347 PRVVAESPVGKAVDVLIVRKGVETVYKVLGRLEDGEKLASGGNGNTDQDKGNKAPAV-- 404
Qy 412 QSGTFSVESAGITLQTHDSSGKHLV-----VVRVSDAERAGLRHGDILAV 459
Db 405 --STASV--LGMTVGLNDETTRKFKSIAADVSGVITDVAKDAAER-GIOPGEVITEI 459
Qy 460 R----ASPR 464
Db 460 AQESVATPK 468
```

```
RESULT 10
ID Q44652 PRELIMINARY; PRT; 500 AA.
AC Q44652;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE IMMUNOREACTIVE STRESS RESPONSE PROTEIN PRECURSOR.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=235;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=2308;
RX MEDLINE=941156447; PubMed=8112833;
RA Roop R.M. II., Fletcher T.W., Sritanganathan N.M., Boyle S.M.,
RA Schurig G.G.;
RT "Identification of an immunoreactive Brucella abortus HtrA stress
RT response protein homolog.";
RL Infect. Immun. 62:1000-1007(1994).
DR EMBL; L09274; AAA53693.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 500 IMMUNOREACTIVE STRESS RESPONSE PROTEIN.
SQ SEQUENCE 500 AA; 51699 MW; 25FE06592D3BD6A6 CRC64;
```

Query Match 28.2%; Score 664.5; DB 2; Length 500;  
Best Local Similarity 36.7%; Pred. No. 1.7e-33;  
Matches 176; Conservative 80; Mismatches 170; Indels 53; Gaps 13;  
Qy 9 ALAALCAALLAGCEKAGSFEGADKKEASFVERIEHTKDDGSGVSMLLP---DFAQLVQSEG 65

Db 14 AVALSASL-----ACAFVVTGTLGALNEARAE-----AVHVTPPQAGFADLVKVR 61

Qy 66 PAVVNIQAAPRTQSGN---AETSDPLADSDPFYEFF-----KRLV 107

Db 62 PAVVSVRVKKDVQTSNRGQFFGPPGFDQLPDGHPPLKREFRDFGMEPRGDSRNRKG 121

Qy 108 PNMEPIFOEADDGGLNFGSGFIISKNGYILTNTHVAVMGSIKVLNDKREYTAKLIGS 167

Db 122 ANKPRGHERP-----VAQSGSFVSDGYVYNNHVVSDGDATVVLDDGTDLDAKLIGA 177

Qy 168 DVQSDVALLKIDATEELPVVK-IGNPKNLKPGEWAAIGAPFGFDSNVTAAGISAKGRSL 226

Db 178 DPTDLAVLKINAPKRFVYSLATIIRCDVGDVWVAVGNDFGLGTVTSIGVSARGADI 237

Qy 227 PNEYTPFIQTDVAINTPNSGGPLFNKLGQVVGINSQIYSRSGFGMGSFAIPIDVAMNV 286

Db 238 GAGPYDDFTQIDAADVKNKSGGPAFDLSGEVIGINTAIFSPSGGTGVIAGFAIPSTAKOV 297

Qy 287 AQLANTKVGORGLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346

Db 298 VDOLIKKSGVEKGLVQIQPVTKDIAASLGLAEEKGAIVASPDQDGAAPAKAGIKAGDVI 357

Qy 347 LSLDGEIRSSGDLPMVGAITPGKEVSLGVWR--KGEIYIK-AKLGNAAEHTGASSKT 403

Db 358 TAVNGETVQDPLARKVANIAPGEAKALTVMKNAEEINTIAAMPNDKKGSGSQSND 417

Qy 404 DEAPYTEQSCIFSVESAGITLQTHDSSGKHLVVRV---SDAAERAGLRHGDEILAV 459

Db 418 NDGGQGE-----TLDSYGLTVVPSD--GKGVVTVDPDSDAADR-GIRSGDIVSV 467

RESULT 11

Q9FD11 PRELIMINARY; PRT; 453 AA.

AC Q9FD11; 1997 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE HTRA-LIKE SERINE PROTEASE.

GN PRS1.

OS Aeromonas hydrophila.

OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;

OC Aeromonas.

OX NCBI\_TaxID=644;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CKH-29;

RA Lin T.-N., Lin T.-J., Liou C.-M.;

RT "Aeromonas hydrophila strain CKH-29 prts1 gene complete cds."

Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

EMBL; AF293977; AAG03073.1; -

DR MEROPS; S01.274; -

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001940; Protease2C.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00595; PDZ; 2.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00834; PROTEASES2C.

DR SMART; SM00228; PDZ; 2.

DR PROSITE; PS50106; PDZ; 2.

DR PROSITE; PS50240; TRYPsin\_DOM; 1.

KW Hydrolase; Protease; Serine protease.

SQ SEQUENCE 453 AA; 47828 MW; B055CE2B844E5148 CRC64;

Query Match 28.0%; Score 660; DB 2; Length 453;

Best Local Similarity 39.2%; Pred. No. 2.8e-33;

Matches 160; Conservative 76; Mismatches 140; Indels 32; Gaps 8;

Qy 114 POEADDGGLNFGSGFII-SKNGYILTNTHVAVMGSIKVLNDKREYTAKLIGSDVQSD 172

Db 82 -EQVSEQFQALSGSVIIDAKKGYVITNAHVHVEADEIKVNLKDGREYAAKKIGEDKQSD 140

Qy 173 VALLKIDATEELPVVKIGNPKNLKPGEWAAIGAPFGFDSNVTAAGISAKGRS-LPNESY 231

Db 141 IALLQIKA-EDLVQIKFADSDDELVRGDYALATGNPGLQTVTSGIVSALGRSGNIENL 199

Qy 232 TPIQTDVAINTPNSGGPLFNKLGQVVGINSQIYSRSGFGMGSFAIPIDVAMNVAEQLK 291

Db 200 ENFIQDAAINSGSGALLNLRLGELIGINTAILPGNNGNIGIGFAIPSPMVRDLSEQIV 259

Qy 292 NTCKVORGOLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVSLDG 351

Db 260 KYGEVRGQGLIGITELTSEVAKTFGYNKDGAFVNVQVPSDAAAKAGIKPGDIIVSIDG 319

Qy 352 GEIRSSGDLPMVGAITPGKEVSLGVWRKGEIYIKAKLGNAAEHTGASSKTDEAPYTEQ 411

Db 320 KAIRSFGEELRAKATMGADKQVALGRLDGEQTVKVTLLKADDSSEILASALHPA----- 374

Qy 412 QSGTFSVESAGITLQTHDSSGKHLVVRVSDAAERAGLRHGDEILAV 459

Db 375 -----LEGAKLS--TTSEPVSGVAVSEIDPRSPAASGLQKGDVVIIGV 415

RESULT 12

P74978 PRELIMINARY; PRT; 478 AA.

AC P74978; 1997 (Tremblrel. 02, Created)

DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE GSRA PROTEIN.

GN GSRA.

OS Yersinia enterocolitica.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Yersinia.

OX NCBI\_TaxID=630;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WA314;

RA MEDLINE=96333330; PubMed=8757824;

RT Yamamoto T., Hanawa T., Ogata S., Kamiya S.;

RT "Identification and characterization of the Yersinia enterocolitica

gsrA gene, which protectively responds to intracellular stress induced

by macrophage phagocytosis and to extracellular environmental stress.";

RL Infect. Immun. 64:2980-2987(1996).

DR EMBL; D78376; BAA11382.1; -

DR MEROPS; S01.273; -

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001940; Protease2C.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00595; PDZ; 2.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00834; PROTEASES2C.

DR SMART; SM00228; PDZ; 2.

DR PROSITE; PS50106; PDZ; 2.

DR PROSITE; PS50240; TRYPsin\_DOM; 1.

KW Hydrolase; Serine protease.

SQ SEQUENCE 478 AA; 49459 MW; E0D43CDA18BDE17C CRC64;

Query Match 27.2%; Score 639.5; DB 2; Length 478;

Best Local Similarity 35.3%; Pred. No. 5.6e-32;

Matches 154; Conservative 79; Mismatches 162; Indels 41; Gaps 8;

Qy 49 SVSMLLPDFAOLVQSEGPVAVVNIQAAPAPRTQNGSGNAE-----TSDSPLADSDPFYE 101

Db 33 SSSQOLPLSLAPMLEKVMPSVSVINVEGSAQVSNAGGIPQPFQFFGDDSPFCQDGSPP-- 90

Qy 102 FFKRLVPMNPETPEAEADDGGL-----NFGSGFII-SKNGYILTNTHVAVMGSIKVL 153

Db 91 -----QGSPPMCQGGPGSGRGAPSKQDFRALGSGVVIIDAKKGYVYNNHVDNATKINV 144

Query Match 28.0%; Score 660; DB 2; Length 453;

Best Local Similarity 39.2%; Pred. No. 2.8e-33;

Matches 160; Conservative 76; Mismatches 140; Indels 32; Gaps 8;

	Qy	291	KNTKGVORGOLGVTIOEVSYGCLAOSFCLDKASACALTALILPGSPAERAGLOAGDVLVSLD	350
	Db	290	KDGKVSGRGLCVTIQDLTDEISGLELKNTRGVLUAKAVOEDCPGDRAKITGDIIIEFA	349
	Qy	351	GGEIRSSGDDLPMVGVAITPCKEVSLGVWRKEELITKAICLNAAEHIT--GASSKTDEAPY	408
	Db	350	DIPVKTKKLURVIADIAPIODEVKVTLRDKLELPKITTSNEEVTIKDSTEETNKKEI	409
	Qy	409	TEQGSGTFVESAGITL-----QTHTDSSCKHLVVVRVSAAAARAGLRHCHDEI	456
	Db	410	TNKENNLSTTKNNITFGNLTEELROKYTIPODKMGIVITNID-EEESSFKIGDLI	464

RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRALINE-NTHI 12;
RX	MEDLINE=98147697; PubMed=9488373;
RA	Loosmore S.M., Yang Y.P., Oomen R., Shortreed J.M., Coleman D.C.,
RT	Klein M.H.;
RR	"The Haemophilus influenzae HtrA protein is a protective antigen.";
RL	Infect. Immun. 66:899-906(1998).
DR	ENBL; AF018151; AAC38202.1; "
DR	MEROPS: S01_274; "
DR	InterPro; IPR001314; Chymotrypsin.
DR	InterPro; IPR001478; PDZ.
DR	InterPro; IPR001940; ProteaseS2C.
DR	InterPro; IPR001254; Trypsin.
DR	Pfam; PF00595; PDZ; 2.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	PRINTS; PR00834; PROTEASES2C.
DR	SMART; SM00228; PDZ; 2.
DR	PROSITE; PS50106; PDZ; 2.
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.
KW	Hydrolase; Serine protease.
FT	NON_TER 1
SQ	SEQUENCE 437 AA; 46376 MW; 27AE2E9F27BE5F6C CRC64;

[illegible]

Db 313 LKAGDVIVAMGOKISSFAEMRAKATATSGAGKEIALTYLRDCK-----THOTKV 361  
QY 400 SSKTDEAPYTEOOSGTFVSAGITLQHTDSSGKHLVVVRVS--DAAERAGLRHGDEIL 457  
Db 362 TLOSDD--QTQADASHLLPALAGAEMSNHDKGVKGLITSVTPKSLAEOGLKKGDVII 419  
QY 458 AV 459  
Db 420 GV 421

Search completed: June 10, 2002, 12:13:55  
Job time: 239 sec

Db 238 VQOILEFCQVRGLGIGKGELNADLAKAFNVSAQQAFVSEVLPKSAEKAGLKAGDII 297  
QY 347 LSLDGEIRSSGDLPLVMVGAITPCGEVSLGVWRKGEIITIKAKLGNAAEHTGASSKTDEA 406  
Db 298 TAMNQKTISSFAEIRAKIATTGAGKELISLYLRDCKSHDYKMKL-QADDSQLSSKT-EL 355  
QY 407 PYTEOOSGTFVSAGITLQHTDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
Db 356 PALD-----GATLKDYDAKGVKGIETIKIPNSLAAQR-GLKSGDIIIGI 399

RESULT 15  
Q9CMS7 PRELIMINARY; PRT; 459 AA.  
AC Q9CMS7;  
DT 01-JUN-2001 (TremBLrel. 17, Created)  
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)  
DE HTRA.  
GN HTRA OR PM0734.  
SS Pasteurella multocida.  
SS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
SS Pasteurella.  
CC NCBI\_TaxID=747;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=PM70;  
RC MEDLINE=21145866; PubMed=11248100;  
RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RA "Complete genomic sequence of Pasteurella multocida Pm70.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
RL -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
CC EMBL: AE006110; AAK02818.1;  
DR MEROPS: S01.274;  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001940; Protease2C.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00595; PDZ; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00834; PROTEASES2C.  
DR SMART: SM00228; PDZ; 2.  
DR SMART: SM00020; Tryp\_Spc; 1.  
DR PROSITE: PS50106; PDZ; 2.  
DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
KW Complete proteome; Hydrolase; Serine protease.  
KW SEQUENCE 459 AA; 48533 MW; 9889E3945812F0FA CRC64;

Query Match 27.1%; Score 637; DB 16; Length 459;  
Best Local Similarity 35.8%; Pred. No. 7.6e-32;  
Matches 151; Conservative 85; Mismatches 150; Indels 36; Gaps 9;  
QY 44 TKDDGSVMLLPDPFALQVSEGPVAVVNIQAAPRTQNGSNAET-DSDPLADSDPFYEF 102  
Db 30 TVDVG--QOVPSLAPMLEKVLPAVVSIV-----EGKAKNHGSTPFQGIPEERF 77  
QY 103 FKRLVPMPIPOEADGGNFGSGFII-SKNGYILTNTVVVAGMGSIKVLINDKREYT 161  
Db 78 FFG-----PDLFNDRAPRNFRTGIGSGVIIINAKEYVLTNNHVIDSADKITVQLQDGRSL 132  
QY 162 AKLIGSDVQSDVALLKDATEELPVYKIGNPKLKGEEVAAIACAPFGDNSVTAGIVSA 221  
Db 133 AKVIGTDELSDVALIQIEKPKNLTAJLTDSDKLRVGDFTVAIGNPFLGQTVTSGIVSA 192  
QY 222 KGRSL--PNESYTPFIQTDVAINPGNSGGPLFNLKGQVGINSGIYSRSGFMGISFAIP 279  
Db 193 LGRSTGSDNGYENYIQTDAVNRNGSGGPLINLQGLIGINTAIISPSGGNAGIAFAIP 252  
QY 280 IDVAMNVAEQKNTGVQGOGLVYIIQVSVGLAQSGLDKASGALIAKILPGSPAERAG 339  
Db 253 SNWANNLVQIILEFGEVRRGMLGKGGELNADLAKAFDIEAQQAQAFVSEVLPNSAAEKAG 312  
QY 340 LOAGDIVLSLDGGEIRSSGDLPLVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGA 399

